GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 18, 2005, 20:14:54 ; Search time 192 Seconds
(without alignments)
524.051 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1198	100.0	229	4	AAM40223	Aam40223 Human pol
2	1198	100.0	229	5	ABG34856	Abg34856 Human can
3	1198	100.0	231	4	AAM42009	Aam42009 Human pol
4	138.5	11.6	164	5	ABG34855	Abg34855 Human cDN
5	113	9.4	233	4	AAM93668	Aam93668 Human pol
6	113	9.4	233	4	AAG93290	Aag93290 Human pro
7	113	9.4	233	7	ADJ69163	Adj69163 Human hea
8	113	9.4	233	8	ADL31517	Adl31517 Human pro
9	92	7.7	695	5	ABB54167	Abb54167 Lactococc
10	92	7.7	695	8	ADS29356	Ads29356 Bacterial
11	90	7.5	269	4	AAO08972	Aao08972 Human pol
12	89	7.4	572	8	ABM83354	Abm83354 Human dia
13	89	7.4	623	4	AAB93182	Aab93182 Human pro
14	89	7.4	623	5	ABB97233	Abb97233 Novel hum
15	89	7.4	652	4	AAB93168	Aab93168 Human pro
16	89	7.4	664	4	AAB83843	Aab83843 Amino aci
17	89	7.4	664	4	AAB20219	Aab20219 Human Chf
18	88.5	7.4	388	8	ADQ08702	Adq08702 Ciona int
19	87.5	7.3	1245	7	ADC32818	Adc32818 Human nov
20	86.5	7.2	470	8	ABM83305	Abm83305 Human dia
21	86.5	7.2	669	6	ABU23878	Abu23878 Protein e
22	86	7.2	158	7	ADH88711	Adh88711 Enterococ
23	85.5	7.1	571	8	ADU07910	Adu07910 Amino aci
24	84.5	7.1	485	5	ABP51336	Abp51336 Human MDD
25	84	7.0	496	4	AAB95788	Aab95788 Human pro
26	84	7.0	496	7	ADC31196	Adc31196 Human nov
27	83.5	7.0	479	4	AAU39678	Aau39678 Propionib
28	83.5	7.0	479	6	ABM36197	Abm36197 Propionib
29	83	6.9	598	2	AAW59461	Aaw59461 Microbial
30	83	6.9	598	2	AAW59463	Aaw59463 Microbial
31	83	6.9	598	2	AAW59457	Aaw59457 Microbial
32	83	6.9	598	2	AAW59458	Aaw59458 Microbial
33	83	6.9	598	2	AAW59459	Aaw59459 Microbial
34	83	6.9	598	2	AAW59462	Aaw59462 Microbial
35	83	6.9	598	2	AAW59460	Aaw59460 Microbial
36	83	6.9	598	2	AAW59464	Aaw59464 Microbial
37	83	6.9	599	2	AAW59496	Aaw59496 Microbial
38	83	6.9	606	1	AAP90110	Aap90110 Polypepti
39	83	6.9	607	2	AAW59456	Aaw59456 Microbial
40	83	6.9	607	2	AAW59465	Aaw59465 Microbial
41	83	6.9	663	6	ABU23749	Abu23749 Protein e
42	82.5	6.9	278	8	ADX72944	Adx72944 Plant ful
43	82.5	6.9	325	3	AAY90242	Aay90242 Human cyc
44	82.5	6.9	493	8	ADN72405	Adn72405 Thale cre
45	82.5	6.9	493	9	AEA16977	Aea16977 Arabidops

ALIGNMENTS

RESULT 1
AAM40223
ID AAM40223 standard; protein; 229 AA.
XX
AC AAM40223;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3368.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59379.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX

PS Example 5; SEQ ID NO 3368; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 229 AA;

Query Match 100.0%; Score 1198; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.7e-127;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

RESULT 2
ABG34856
ID ABG34856 standard; protein; 229 AA.
XX
AC ABG34856;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cancer related protein encoded by cDNA 85P1B3.
XX
KW Human; cytostatic; 85P1B3; cancer; immunogen; chromosome 15q14.
XX
OS Homo sapiens.
XX
PN WO200218578-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-US026838.
XX
PR 28-AUG-2000; 2000US-0228432P.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Raitano AB, Faris M, Hubert RS, Afar D, Ge W, Challita-Eid P;
PI Jakobovits A;
XX
DR WPI; 2002-382963/41.
DR N-PSDB; ABK70506.
XX
PT Composition for modulating the status of 85P1B3 protein or a molecule
PT comprising a substance e.g. antibody specific to, nucleic acid encoding,
PT or ribozyme of 85P1B3.
XX
PS Claim 34; Fig 2; 201pp; English.
XX
CC The invention relates to a composition comprising a substance that
CC modulate the status of 85P1B3, where the status of a cell expresses
CC 85P1B3 gene product is modulated. Also included are a composition
CC comprising a peptide region of 5 amino acids of the 85P1B3 protein, in
CC any whole number increment up to 229 that includes an aa position
CC selected from an aa position having a value greater than 0.5 in the
CC Hydrophilicity profile, an aa position having a value less than 0.5 in
CC the hydropathicity profile, an aa position having a value greater than
CC 0.5 in the percent accessible residue profile, an aa position having a
CC value greater than 0.5 in the average flexibility profile, or an aa
CC position having a value greater than 0.5 in the beta-turn profile; a
CC polynucleotide that encodes analogue peptide of 8, 9, 10 or 11 contiguous
CC residues of the 85P1B3 protein; a recombinant protein comprising the
CC antigen-binding region of a monoclonal antibody; a non-human transgenic
CC animal that produces an antibody that binds to the 85P1B3 protein; a
CC hybridoma that produces antibody specific to the protein; a single chain
CC monoclonal antibody (MAb) that comprises the variable domains of the
CC heavy and monoclonal antibodies specific to the protein; a vector
CC comprising a polynucleotide that encodes the MAb; inhibiting growth of
CC cancer cells or treating a patient who bears cancer cells that expresses
CC the protein, by administering the protein, antibody, polynucleotide
CC encoding the protein, antisense polynucleotide to the polynucleotide,
CC ribozyme that cleaves the polynucleotide and T cells that specifically
CC recognize the protein; and generating a mammalian immune response
CC directed to the protein exposing cells of the mammal's immune system to
CC an immunogenic portion of the protein or polynucleotide. The composition,
CC which comprises an antibody specific to the protein, is useful for
CC delivering a cytotoxic agent to a cell that expresses the protein by
CC providing a cytotoxic agent conjugated to antibody and exposing the cell
CC to the antibody-agent conjugate. The methods are useful for inhibiting
CC growth of cancer cells or treating a patient who bears cancer cells that
CC expresses the protein, for generating a mammalian immune response
CC directed to the protein, for detecting the presence of the protein or
CC polynucleotide in a biological sample in a patient who has or who is
CC suspected of having cancer and for monitoring 85P1B3 in a biological
CC sample from a patient who has or who is suspected of having cancer. The
CC gene for 85P1B3 is located on human chromosome 15q14. The present
CC sequence is the 85P1B3 protein
XX
SQ Sequence 229 AA;

Query Match 100.0%; Score 1198; DB 5; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.7e-127;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

RESULT 3
AAM42009
ID AAM42009 standard; protein; 231 AA.
XX
AC AAM42009;
XX
DT 22-OCT-2001 (first entry)

XX
DE Human polypeptide SEQ ID NO 6940.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI61165.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 2; SEQ ID NO 6940; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 231 AA;

Query Match 100.0%; Score 1198; DB 4; Length 231;
Best Local Similarity 100.0%; Pred. No. 9.8e-127;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 62

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 122

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 182

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 231
```

RESULT 4
ABG34855
ID ABG34855 standard; protein; 164 AA.
XX
AC ABG34855;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cDNA 85P1B3 splice variant, open reading frame #3.
XX
KW Human; cytostatic; 85P1B3; cancer; immunogen; chromosome 15q14.
XX
OS Homo sapiens.
XX
PN WO200218578-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-US026838.
XX
PR 28-AUG-2000; 2000US-0228432P.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Raitano AB, Faris M, Hubert RS, Afar D, Ge W, Challita-Eid P;
PI Jakobovits A;
XX
DR WPI; 2002-382963/41.
DR N-PSDB; ABK70504.
XX
PT Composition for modulating the status of 85P1B3 protein or a molecule
PT comprising a substance e.g. antibody specific to, nucleic acid encoding,
PT or ribozyme of 85P1B3.
XX
PS Example 38; Page 124; 201pp; English.
XX
CC The invention relates to a composition comprising a substance that
CC modulate the status of 85P1B3, where the status of a cell expresses
CC 85P1B3 gene product is modulated. Also included are a composition
CC comprising a peptide region of 5 amino acids of the 85P1B3 protein, in
CC any whole number increment up to 229 that includes an aa position
CC selected from an aa position having a value greater than 0.5 in the
CC Hydrophilicity profile, an aa position having a value less than 0.5 in
CC the hydropathicity profile, an aa position having a value greater than
CC 0.5 in the percent accessible residue profile, an aa position having a
CC value greater than 0.5 in the average flexibility profile, or an aa
CC position having a value greater than 0.5 in the beta-turn profile; a
CC polynucleotide that encodes analogue peptide of 8, 9, 10 or 11 contiguous
CC residues of the 85P1B3 protein; a recombinant protein comprising the
CC antigen-binding region of a monoclonal antibody; a non-human transgenic
CC animal that produces an antibody that binds to the 85P1B3 protein; a
CC hybridoma that produces antibody specific to the protein; a single chain
CC monoclonal antibody (MAb) that comprises the variable domains of the
CC heavy and monoclonal antibodies specific to the protein; a vector
CC comprising a polynucleotide that encodes the MAb; inhibiting growth of
CC cancer cells or treating a patient who bears cancer cells that expresses
CC the protein, by administering the protein, antibody, polynucleotide
CC encoding the protein, antisense polynucleotide to the polynucleotide,
CC ribozyme that cleaves the polynucleotide and T cells that specifically
CC recognize the protein; and generating a mammalian immune response
CC directed to the protein exposing cells of the mammal's immune system to
CC an immunogenic portion of the protein or polynucleotide. The composition,
CC which comprises an antibody specific to the protein, is useful for
CC delivering a cytotoxic agent to a cell that expresses the protein by
CC providing a cytotoxic agent conjugated to antibody and exposing the cell

```
CC   to the antibody-agent conjugate. The methods are useful for inhibiting
CC   growth of cancer cells or treating a patient who bears cancer cells that
CC   expresses the protein, for generating a mammalian immune response
CC   directed to the protein, for detecting the presence of the protein or
CC   polynucleotide in a biological sample in a patient who has or who is
CC   suspected of having cancer and for monitoring 85P1B3 in a biological
CC   sample from a patient who has or who is suspected of having cancer. The
CC   gene for 85P1B3 is located on human chromosome 15q14. The present
CC   sequence is the 85P1B3 protein
XX
SQ   Sequence 164 AA;

  Query Match             11.6%;  Score 138.5;  DB 5;  Length 164;
  Best Local Similarity   82.1%;  Pred. No. 8.6e-07;
  Matches   32;  Conservative    1;  Mismatches    1;  Indels    5;  Gaps    1;

Qy        162 FCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              | |||     |||||||||||||||||||||||||||:|
Db          2 FFLSS-----YLLKTKAIVNASEMDIQNVPLSEKIAEVK 35


RESULT 5
AAM93668
ID   AAM93668 standard; protein; 233 AA.
XX
AC   AAM93668;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3550.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94603.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3550; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence is a polypeptide encoded by a full length
CC   human cDNA of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in CD-ROM
CC   format directly from EPO
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 4;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0011;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::   :         |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 6
AAG93290
ID   AAG93290 standard; protein; 233 AA.
XX
AC   AAG93290;
XX
DT   13-SEP-2001  (first entry)
XX
DE   Human protein HP10650.
XX
KW   Human; gene therapy; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200142302-A1.
XX
PD   14-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-JP008631.
XX
PR   06-DEC-1999;   99JP-00346863.
PR   06-DEC-1999;   99JP-00346864.
PR   08-FEB-2000; 2000JP-00031062.
PR   10-FEB-2000; 2000JP-00034090.
PR   10-FEB-2000; 2000JP-00034091.
PR   14-FEB-2000; 2000JP-00035829.
PR   14-FEB-2000; 2000JP-00035899.
PR   14-MAR-2000; 2000JP-00071161.
PR   30-MAY-2000; 2000JP-00160851.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Kato S,  Eguchi C,  Saeki M;
XX
DR   WPI; 2001-381646/40.
DR   N-PSDB; AAH68575.
XX
PT   Human protein originated from tumor cell line, applicable as drug,
PT   reagent for studying intracellular protein networks and protein source
PT   for drug screening, also encoded cDNA for gene diagnosis and gene
PT   therapy.
XX
PS   Claim 1; Page 303-304; 471pp; Japanese.
XX
CC   The present sequence is a human protein. The human protein, preferably
CC   originated from tumour cell line, is applicable as a drug, a reagent for
CC   studying intracellular protein networks and a protein source for
CC   screening proteins for binding low molecular weight drugs. The human
CC   protein coding sequence is useful for gene diagnosis and gene therapy,
CC   expression vectors and transformant cells for detection of ligands and
CC   receptors
XX
```

```
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 4;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0011;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |      | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 7
ADJ69163
ID   ADJ69163 standard; protein; 233 AA.
XX
AC   ADJ69163;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human heat mitochondrial protein as a therapeutic target SeqID969.
XX
KW   mitochondrial; human; screening assay; diabetes mellitus;
KW   Huntington's disease; osteoarthritis;
KW   Leber's hereditary optic neuropathy; LHON;
KW   mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW   myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW   neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW   osteopathic; ophthalmological; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2003087768-A2.
XX
PD   23-OCT-2003.
XX
PF   04-APR-2003; 2003WO-US010870.
XX
PR   12-APR-2002; 2002US-0372843P.
PR   17-JUN-2002; 2002US-0389987P.
PR   20-SEP-2002; 2002US-0412418P.
XX
PA   (MITO-) MITOKOR.
PA   (BUCK-) BUCK INST AGE RES.
XX
PI   Ghosh SS,  Fahy ED,  Zhang B,  Gibson BW,  Taylor SW,  Glenn GM;
PI   Warnock DE;
XX
DR   WPI; 2003-845369/78.
XX
PT   Identifying a mitochondrial target for drug screening assays and for
PT   treating diseases associated with altered mitochondrial function,
PT   comprises detecting a modified polypeptide in a sample and correlating
PT   with the disease.
XX
PS   Claim 1; SEQ ID NO 969; 180pp; English.
XX
CC   This invention relates to novel mitochondrial targets that can be used
CC   for therapeutic intervention in treating a disease associated with
CC   altered mitochondrial function. Specifically, it refers to a method for
CC   identifying proteins of the human heart mitochondrial proteome that are
CC   useful for drug screening assays, as well as therapeutic targets. The
CC   present invention describes a method for identifying such proteins that
CC   can be used in the treatment of various diseases associated with altered
CC   mitochondrial function including diabetes mellitus, Huntington's disease,
CC   osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC   encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC   ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC   compositions have neuroprotective, nootropic, antidiabetic,
CC   anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC   cytostatic activities. This polypeptide sequence is a human heart
CC   mitochondrial protein of the invention.
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 7;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0011;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |      | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 8
ADL31517
ID   ADL31517 standard; protein; 233 AA.
XX
AC   ADL31517;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human protein encoded by a full length cDNA clone SeqID 3550.
XX
KW   human; medicine; signal transduction; glycoprotein; transcription;
KW   oligo-capping method.
XX
OS   Homo sapiens.
XX
PN   EP1396543-A2.
XX
PD   10-MAR-2004.
XX
PF   07-JUL-2000; 2003EP-00025638.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183865.
PR   07-JUL-2000; 2000EP-00114089.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2004-204755/20.
DR   N-PSDB; ADL31516.
XX
PT   New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT   length human cDNAs.
XX
PS   Example 1; SEQ ID NO 3550; 1340pp; English.
XX
```

```
CC   This invention relates to a novel primers useful for synthesising full
CC   length cDNA molecules that encode human proteins. Specifically, it refers
CC   to secretory or membrane proteins that are potential therapeutic agents/
CC   target molecules in the field of medicine, and in particular genes
CC   encoding proteins that are associated with signal transduction,
CC   glycoproteins and transcription. The present invention describes a method
CC   for efficiently cloning a full length human cDNA from both the 5' and 3'
CC   ends using the oligo-capping method. This polypeptide sequence is a full
CC   length human protein of the invention.
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 8;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0011;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                 |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 9
ABB54167
ID   ABB54167 standard; protein; 695 AA.
XX
AC   ABB54167;
XX
DT   29-AUG-2003  (revised)
DT   16-MAY-2002  (first entry)
XX
DE   Lactococcus lactis protein yihC.
XX
KW   Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS   Lactococcus lactis; IL1403.
XX
PN   FR2807446-A1.
XX
PD   12-OCT-2001.
XX
PF   11-APR-2000; 2000FR-00004630.
XX
PR   11-APR-2000; 2000FR-00004630.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Bolotine A,  Sorokine A,  Renault P,  Ehrlich SD;
XX
DR   WPI; 2002-043418/06.
XX
PT   New nucleotide sequence useful in the identification or Lactococcus
PT   lactis and related species.
XX
PS   Claim 6; SEQ ID NO 869; 2504pp; French.
XX
CC   The present invention is related to a Lactococcus lactis nucleotide
CC   sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
CC   acid sequence is useful in the detection and/or amplification of nucleic
CC   acid sequence, particularly to identify Lactococcus lactis or related
CC   species. The proteins of the invention are useful for the biosynthesis or
CC   biodegradation of a composition of interest. The invention helps research
CC   in lactic bacteria, particularly useful in the production of yogurt and
CC   cheese. Note: The sequence data for this patent is based on equivalent
CC   patent WO200177334 (published 18-OCT-2001) which is available in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 695 AA;

  Query Match             7.7%;  Score 92;  DB 5;  Length 695;
  Best Local Similarity   25.3%;  Pred. No. 1.5;
  Matches   38;  Conservative   28;  Mismatches   50;  Indels   34;  Gaps    7;

Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
              || ::|            ||  | :| ||     ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI------------QNVPLSE-K 195
              |          |    :: :|:: |||: | :   |:              ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
              : :|   :|  : :: :   ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241


RESULT 10
ADS29356
ID   ADS29356 standard; protein; 695 AA.
XX
AC   ADS29356;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #18389.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 18389; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
```

```
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant.growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 695 AA;

  Query Match             7.7%;  Score 92;  DB 8;  Length 695;
  Best Local Similarity   25.3%;  Pred. No. 1.5;
  Matches   38;  Conservative   28;  Mismatches   50;  Indels   34;  Gaps    7;

Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
              || ::|           ||  | :| ||     ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI------------QNVPLSE-K 195
              |          |    :: :|:: |||: | :   |:             ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
              : :|   :|  : :: :   ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241


RESULT 11
AAO08972
ID   AAO08972 standard; protein; 269 AA.
XX
AC   AAO08972;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 22864.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US004927.
XX
PR   28-FEB-2000; 2000US-00515126.
PR   18-MAY-2000; 2000US-00577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI88903.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT   and treating e.g. leukemia, inflammation and immune disorders.
XX
PS   Claim 20; SEQ ID NO 22864; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 269 AA;

  Query Match             7.5%;  Score 90;  DB 4;  Length 269;
  Best Local Similarity   23.0%;  Pred. No. 0.58;
  Matches   52;  Conservative   26;  Mismatches   74;  Indels   74;  Gaps   11;

Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::     ::||  |
Db         39 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 95

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||              ||| |   |    || |  :|
Db         96 PMPDRRAEREQDPRVAPQ-------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 139

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIHGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|                | :
Db        140 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML---------------TES 183

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
              :|  || | ||      | ::||    ::    :    ||:| ||
Db        184 LMALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 229


RESULT 12
ABM83354
ID   ABM83354 standard; protein; 572 AA.
XX
AC   ABM83354;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human diagnostic and therapeutic pprotein SEQ ID NO:3603.
XX
KW   gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS   Homo sapiens.
XX
PN   WO2004023973-A2.
XX
PD   25-MAR-2004.
XX
PF   12-SEP-2003; 2003WO-US028227.
XX
PR   12-SEP-2002; 2002US-0410259P.
PR   12-SEP-2002; 2002US-0410260P.
XX
PA   (INCY-) INCYTE CORP.
XX
PI   Schmidt JP,  Wright RJ,  Bruns CM,  Marjanovic MM,  Shen F;
PI   Harthshorne TA,  Suchorolski MT,  Altus CM,  Pitts SJ,  Elder LV;
PI   Mooney EM,  Delegeane AM,  Panesar IS,  Banville SC,  Reddy TP;
PI   Stevens KA,  Blanchard JL,  Panzer SR,  Wang X,  Au AP,  Gerstin EH;
PI   Peralta CH,  Anderson SB,  Rioux P,  Shen EJ,  Wu MC,  Stuve LL;
```

PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN42006.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 572 AA;

Query Match 7.4%; Score 89; DB 8; Length 572;
Best Local Similarity 23.0%; Pred. No. 2.4;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

Qy 8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
Db 342 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 398

Qy 51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
Db 399 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 442

Qy 100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
Db 443 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 486

Qy 154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
Db 487 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 532

RESULT 13
AAB93182
ID AAB93182 standard; protein; 623 AA.
XX
AC AAB93182;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12128.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12128; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 623 AA;

Query Match 7.4%; Score 89; DB 4; Length 623;
Best Local Similarity 23.0%; Pred. No. 2.7;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

Qy 8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
Db 393 YRRQAAQPPH---CPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQ 449

Qy 51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
Db 450 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 493

Qy 100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
Db 494 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 537

Qy 154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
Db 538 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYQQNIPASE 583

RESULT 14
ABB97233
ID ABB97233 standard; protein; 623 AA.
XX
AC ABB97233;
XX

DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 501.
XX
KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR N-PSDB; ABN32419.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Example 2; SEQ ID NO 501; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 623 AA;

Query Match 7.4%; Score 89; DB 5; Length 623;
Best Local Similarity 23.0%; Pred. No. 2.7;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

Qy 8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
Db 393 YRRQAAQPPH---CPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQ 449

Qy 51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
Db 450 PMPDRRAERERQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 493

Qy 100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
Db 494 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML---------------TES 537

Qy 154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
Db 538 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYQQNIPASE 583

RESULT 15
AAB93168
ID AAB93168 standard; protein; 652 AA.
XX
AC AAB93168;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12100.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12100; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 652 AA;

Query Match 7.4%; Score 89; DB 4; Length 652;
Best Local Similarity 23.0%; Pred. No. 2.9;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

Qy 8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
Db 422 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 478

Qy 51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
Db 479 PMPDRRAERERQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 522

Qy 100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
Db 523 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML---------------TES 566

```
Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
                :|  || | ||      | ::||    ::    :    ||:| ||
Db        567 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 612
```

Search completed: November 18, 2005, 20:31:06
Job time : 196 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 18, 2005, 20:24:56 ; Search time 48 Seconds
(without alignments)
394.432 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	7.2	158	2	US-09-134-000C-6596	Sequence 6596, Ap
2	83	6.9	457	6	5268463-7	Patent No. 5268463
3	81.5	6.8	456	6	5432081-7	Patent No. 5432081
4	81.5	6.8	457	1	US-08-882-704A-6	Sequence 6, Appli
5	81.5	6.8	457	2	US-09-151-957-6	Sequence 6, Appli
6	80.5	6.7	1711	1	US-08-342-930-2	Sequence 2, Appli
7	78.5	6.6	325	1	US-09-018-576-3	Sequence 3, Appli
8	78.5	6.6	325	1	US-09-018-576-12	Sequence 12, Appl
9	78.5	6.6	325	2	US-09-248-137-3	Sequence 3, Appli
10	78.5	6.6	325	2	US-09-248-137-12	Sequence 12, Appl
11	78.5	6.6	452	2	US-09-771-161A-237	Sequence 237, App
12	78	6.5	190	2	US-09-252-991A-19049	Sequence 19049, A
13	77	6.4	607	2	US-09-537-682-1	Sequence 1, Appli
14	77	6.4	719	1	US-08-520-933-3	Sequence 3, Appli
15	77	6.4	719	2	US-09-285-040-3	Sequence 3, Appli
16	77	6.4	738	6	5262177-2	Patent No. 5262177
17	77	6.4	776	2	US-09-949-016-10829	Sequence 10829, A
18	76.5	6.4	282	2	US-09-845-713A-2	Sequence 2, Appli
19	76	6.3	351	2	US-09-902-540-13214	Sequence 13214, A
20	76	6.3	1001	2	US-10-104-047-2283	Sequence 2283, Ap
21	75.5	6.3	310	1	US-08-129-456A-36	Sequence 36, Appl
22	75.5	6.3	310	1	US-08-705-868-3	Sequence 3, Appli
23	75.5	6.3	310	2	US-09-123-615-3	Sequence 3, Appli
24	75.5	6.3	310	2	US-08-360-821B-35	Sequence 35, Appl
25	75.5	6.3	310	2	US-09-851-026-35	Sequence 35, Appl
26	75.5	6.3	1449	2	US-08-840-062-6	Sequence 6, Appli
27	75.5	6.3	1723	2	US-09-194-612A-31	Sequence 31, Appl
28	75.5	6.3	4545	1	US-08-804-227C-14	Sequence 14, Appl
29	75.5	6.3	4550	1	US-08-804-227C-8	Sequence 8, Appli
30	75.5	6.3	4550	1	US-08-804-198-2	Sequence 2, Appli
31	75	6.3	78	2	US-09-270-767-43496	Sequence 43496, A
32	75	6.3	458	2	US-09-800-170-3	Sequence 3, Appli
33	75	6.3	717	6	5262177-5	Patent No. 5262177
34	75	6.3	784	2	US-09-724-586A-1	Sequence 1, Appli
35	75	6.3	784	2	US-09-600-823-1	Sequence 1, Appli
36	75	6.3	784	2	US-09-724-666-1	Sequence 1, Appli
37	75	6.3	1088	2	US-09-920-804-2	Sequence 2, Appli
38	75	6.3	1088	2	US-09-920-804-4	Sequence 4, Appli
39	75	6.3	1129	2	US-09-734-674-2	Sequence 2, Appli
40	75	6.3	1129	2	US-10-274-990-2	Sequence 2, Appli
41	75	6.3	2227	2	US-08-475-886-4	Sequence 4, Appli
42	75	6.3	2227	2	US-09-653-499-4	Sequence 4, Appli
43	75	6.3	2227	2	US-10-135-988-4	Sequence 4, Appli
44	74.5	6.2	211	2	US-09-949-016-11409	Sequence 11409, A
45	74.5	6.2	279	2	US-09-107-433-3335	Sequence 3335, Ap

ALIGNMENTS

RESULT 1
US-09-134-000C-6596
; Sequence 6596, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6596
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6596

Query Match 7.2%; Score 86; DB 2; Length 158;
Best Local Similarity 26.5%; Pred. No. 0.18;
Matches 40; Conservative 27; Mismatdhes 54; Indels 30; Gaps 9;

```
Qy         87 LADSVHLAWDLSRSLGAVV---FSRVTNNVVLEA--PFLVGIEGSLKG--------STYN 133
              | : ::|| :  | |::     |: | |:: :|   |:   ||   |          | |
Db         17 LTEYIYLASEQPLSTGSITLNGLSKPTGNLLNKAFFFFINDFEGDPPGKFSYSNYLKTLN 76

Qy        134 LLFCGSCGIPVGFHLYSTHAALAAL---RGHFCLSSDKMVCYLLKTKAIVNASEMDIQNV 190
              |  | :|       | :|:| |   | ||     |  |    ||  : ::    :|
Db         77 YQVCSMCSLPSDNQNIETPSAIALLGYIRNHF-----KNTC---TTKLYILSALNGFENK 128

Qy        191 PLSE----KIAELKEKIVLTHNRLKSLMKIL 217
                |:|:    |:::|  |  | :  :| |:||:
Db        129 PISKTVTRKLSDLSYK-DLYYQEMK-LLKII 157
```

RESULT 2
5268463-7
;Patent No. 5268463
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: PLAANT PROMOTER a-GLUCURONIDASE GENE
;CONSTRUCT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,976
; FILING DATE: 08-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 119,102

```
;      FILING DATE: 10-NOV-1987
;      APPLICATION NUMBER: 264,586
;      FILING DATE: 31-OCT-1988
;SEQ ID NO:7:
;      LENGTH: 457
5268463-7

  Query Match             6.9%;  Score 83;  DB 6;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 2;
  Matches   37;  Conservative   30;  Mismatches   54;  Indels   26;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        316 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 373

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        374 QAIGGS--SIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKETVVEIDNRKKVQQQLISDIT 456


RESULT 3
5432081-7
;Patent No. 5432081
;    APPLICANT: JEFFERSON, RICHARD A.
;    TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E.COLI
;GLUCORONIDE PERMEASE GENE
;    NUMBER OF SEQUENCES: 10
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/138,546
;      FILING DATE: 15-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 447,976
;      FILING DATE: 08-DEC-1989
;      APPLICATION NUMBER: 264,586
;      FILING DATE: 31-OCT-1988
;      APPLICATION NUMBER: 119,102
;      FILING DATE: 10-NOV-1987
;SEQ ID NO:7:
;      LENGTH: 456
5432081-7

  Query Match             6.8%;  Score 81.5;  DB 6;  Length 456;
  Best Local Similarity   25.2%;  Pred. No. 2.9;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        316 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 373

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        374 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 430

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        431 K--KFKEIVVEIDNRKKVQQQLISDIT 455


RESULT 4
US-08-882-704A-6
; Sequence 6, Application US/08882704A
; Patent No. 5879906
;  GENERAL INFORMATION:
;    APPLICANT:  Jefferson, Richard A.
;    APPLICANT:  Wilson, Katherine J.
;    APPLICANT:  Leader, Michael
;    TITLE OF INVENTION:  GLUCURONIDE REPRESSORS AND USES THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/882,704A
;      FILING DATE:  25-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5879906tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  190106.404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-882-704A-6

  Query Match             6.8%;  Score 81.5;  DB 1;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 2.9;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456


RESULT 5
US-09-151-957-6
; Sequence 6, Application US/09151957
; Patent No. 6429292
;  GENERAL INFORMATION:
;       APPLICANT: Jefferson, Richard A.
;                  Wilson, Katherine J.
;                  Leader, Michael
;       TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF
;       NUMBER OF SEQUENCES: 19
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: SEED and BERRY LLP
;            STREET: 6300 Columbia Center, 701 Fifth Avenue
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/151,957
;            FILING DATE: 11-Sep-1998
;            CLASSIFICATION: <Unknown>
```

```
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 08/882,704
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: No. 6429292tenburg Ph.D., Carol
;             REGISTRATION NUMBER: 39,317
;             REFERENCE/DOCKET NUMBER: 190106.404
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 457 amino acids
;             TYPE: amino acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-151-957-6

  Query Match             6.8%;  Score 81.5;  DB 2;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 2.9;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |      ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456


RESULT 6
US-08-342-930-2
; Sequence 2, Application US/08342930
; Patent No. 5821084
;  GENERAL INFORMATION:
;    APPLICANT:  OLMSTED, ELIZABETH A.
;    APPLICANT:  MAURO, LAURA J.
;    APPLICANT:  DAVIS, ALAN R.
;    APPLICANT:  DIXON, JACK E.
;    TITLE OF INVENTION:  OSTEOBLAST-TESTICULAR PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/342,930
;      FILING DATE:  21-NOV-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20344-20975.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1711 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-342-930-2

  Query Match             6.7%;  Score 80.5;  DB 1;  Length 1711;
  Best Local Similarity   29.6%;  Pred. No. 28;
  Matches   21;  Conservative    8;  Mismatches   17;  Indels   25;  Gaps    1;

Qy         36 SMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAW 95
              ||:  : ||  : | || | ||                            ||| |:|: ||:|
Db        753 SMQAGSAVVNLAWPSGPLGQGA------------------------CHAQLSDAGHLSW 787

Qy         96 DLSRSLGAVVF 106
              :    ||  :|
Db        788 EQPLKLGQELF 798


RESULT 7
US-09-018-576-3
; Sequence 3, Application US/09018576
; Patent No. 5968800
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,576
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-018-576-3

  Query Match             6.6%;  Score 78.5;  DB 1;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 3.8;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG----------------AEEPAAG 63
              |   :|: :      || | | |     : | | |                |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
              |: |:||      |   | |||                              ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA------------------------NNIVHRDLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
```

```
              |:   | || : : |    |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                : :||:|:     |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
              |
Db        237 S 237


RESULT 8
US-09-018-576-12
; Sequence 12, Application US/09018576
; Patent No. 5968800
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,576
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-018-576-12

  Query Match             6.6%;  Score 78.5;  DB 1;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 3.8;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
              |   :|: :      || | | |     : | | |                   |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
              |: |:||     |   | |||                           ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRNLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
              |:   | || : : |.   |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                : :||:|:     |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
              |
Db        237 S 237


RESULT 9
US-09-248-137-3
; Sequence 3, Application US/09248137
; Patent No. 6030788
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/248,137
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/018,576
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-248-137-3

  Query Match             6.6%;  Score 78.5;  DB 2;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 3.8;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
              |   :|: :      || | | |     : | | |                   |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
              |: |:||     |   | |||                           ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRDLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
              |:   | || : : |.   |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                : :||:|:     |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
              |
Db        237 S 237


RESULT 10
```

```
US-09-248-137-12
; Sequence 12, Application US/09248137
; Patent No. 6030788
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/248,137
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/018,576
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-248-137-12

  Query Match             6.6%;  Score 78.5;  DB 2;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 3.8;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
                 |   :|: :      || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
               |: |:||      |   | |||                            ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRNLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
               |:   | || : : |    |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                 : :||:|:      |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
               |
Db        237 S 237


RESULT 11
US-09-771-161A-237
; Sequence 237, Application US/09771161A
; Patent No. 6936450
; GENERAL INFORMATION:
;  APPLICANT: LEVINE, et al.
;  TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
;  FILE REFERENCE: 802620-2005.1
;  CURRENT APPLICATION NUMBER: US/09/771,161A
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 09/724,676
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: 136776
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: 135619
;  PRIOR FILING DATE: 2000-04-12
;  NUMBER OF SEQ ID NOS: 273
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 237
;   LENGTH: 452
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-771-161A-237

  Query Match             6.6%;  Score 78.5;  DB 2;  Length 452;
  Best Local Similarity   23.2%;  Pred. No. 6.3;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
                 |   :|: :      || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
               |: |:||      |   | |||                            ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRDLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
               |:   | || : : |    |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                 : :||:|:      |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
               |
Db        237 S 237


RESULT 12
US-09-252-991A-19049
; Sequence 19049, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 19049
;   LENGTH: 190
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19049

  Query Match             6.5%;  Score 78;  DB 2;  Length 190;
  Best Local Similarity   32.9%;  Pred. No. 1.9;
  Matches   24;  Conservative    6;  Mismatches   29;  Indels   14;  Gaps    3;

Qy          4 QPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAG 63
                :| ||  ||  |  | : |   || |:||              | : : |  ||  |
Db         94 RPCRHTRRCVRPAAGGYPG---RAGDRAS----------VARSTGPGLPAAPGRPRGHC- 139
```

```
Qy          64 PQLPSWLQPERCA 76
               |  | | :  |||
Db         140 PAPPRWPRRRRCA 152


RESULT 13
US-09-537-682-1
; Sequence 1, Application US/09537682
; Patent No. 6303357
; GENERAL INFORMATION:
;  APPLICANT: TAKEUCHI, Kenichi
;  APPLICANT: KOIDE, Yoshinao
;  APPLICANT: NAKANISHI, Yuji
;  APPLICANT: SUZUKI, Satoru
;  TITLE OF INVENTION: L-ALPHA-GLYCEROPHOSPHATE OXIDASE GENE, RECOMBINATNT
;  TITLE OF INVENTION: DNA, AND METHOD FOR PRODUCING MODIFIED
;  TITLE OF INVENTION: L-ALPHA-GLYCEROPHOSPHATE OXIDASE GENE
;  FILE REFERENCE: A20-121814C/KI
;  CURRENT APPLICATION NUMBER: US/09/537,682
;  CURRENT FILING DATE: 2000-03-29
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 607
;   TYPE: PRT
;   ORGANISM: Enterococcus faecium No. 6303357 7044
US-09-537-682-1

  Query Match             6.4%;  Score 77;  DB 2;  Length 607;
  Best Local Similarity   23.2%;  Pred. No. 15;
  Matches   49;  Conservative   30;  Mismatches   76;  Indels   56;  Gaps   10;

Qy          19 DFCGGTERAIDQASFTTSM---------EWDTQVVKGSSPLGP----AGLGAEEPAAGPQ 65
               ||  ||      :| :| :          :| :||  :  :|       |:    |   |
Db          51 DFAEGT------SSRSTKLVHGGIRYLKTFDVEVV--ADTVGERAVVQGIAPHIPKPDPM 102

Qy          66 LPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTN------NVVLEAPF 119
               |    : |    |           || :| ||   |  |  ::  |       |:   ||
Db         103 LLPIYEDEGATTFNMF--------SVKVAMDLYDKLANVTGTKYENYTLTPEEVLEREPF 154

Qy         120 LVGIEGSSLKGSTYNLLFCGSCGIPVGF-----HLYSTHAALAALRGHFCLSSDKMVCYLL 174
               |   :  |||:          |: : |       |   :   ||  | : :|  | | :|
Db         155 LK--KEGLKGA----------GVYLDFRNNDARLVIDNIKKAAEDGAYLVSKMKAVGFLY 202

Qy         175 KTKAIVNASEMDIQNVPLSEKIAELKEKIVL 205
               :   ||     |:    |:::: |:| |:|:
Db         203 EGDQIVGVKARDL----LTDEVIEIKSKLVI 229


RESULT 14
US-08-520-933-3
; Sequence 3, Application US/08520933
; Patent No. 5981194
;  GENERAL INFORMATION:
;    APPLICANT:  Jefferies, Wilfred A.
;    APPLICANT:  McGeer, Patrick L.
;    APPLICANT:  Rothenberger, Sylvia
;    APPLICANT:  Food, Michael R.
;    APPLICANT:  Yamada, Tatsuo
;    APPLICANT:  Kennard, Malcolm
;    TITLE OF INVENTION:  Use of p97 and Iron Binding Proteins
;    TITLE OF INVENTION:  as Diagnostic and Therapeutic Agents
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bereskin & Parr
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/520,933
;      FILING DATE:  August 31, 1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shona S. McDiarmid
;      REGISTRATION NUMBER:  38,798
;      REFERENCE/DOCKET NUMBER:  7685-006
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  416-364-7311
;      TELEFAX:  416-361-1398
;      TELEX:  06-23115
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  719 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-520-933-3

  Query Match             6.4%;  Score 77;  DB 1;  Length 719;
  Best Local Similarity   23.7%;  Pred. No. 19;
  Matches   59;  Conservative   25;  Mismatches   87;  Indels   78;  Gaps   15;

Qy          17 RGDFCGGTERAIDQASFTTSMEW-----------DTQVVKGSSPLGPAGLGAEEPAAGP 64
               |||  |  |   |::      ::           |  || |: |      |   |
Db         171 RGDSSG--EGVCDKSPLERYYDYSGAFRCLAEGAGDVAFVKHSTVL--------ENTDGK 220

Qy          65 QLPSWLQP------------ERCAVFQCAQCH--AVLADSVHLAWDLSRSLGAVVFSRV 109
               |||| |              |  | :  |||   | | :| :  |     | ::| |:
Db         221 TLPSWGQALLSQDFELLCRDGSRADVTEWRQCHLARVPAHAVVVRADTD---GGLIF-RL 276

Qy         110 TNNVVLEAPFLVGIEGS----LKGSTY---NLLFCGSCG--IPVGFHLYST---HAALAA 157
               |    |   |  |||         |   :|||  |    :|:    |      | | |
Db         277 LN----EGQRLFSHEGSSFQMFSSEAYGQKDLLFKDSTSELVPIATQTYEAWLGHEYLHA 332

Qy         158 LRGHFCLSSDKMVCYL----LKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLKSL 213
               ::|  |   :::  ||    | |  |    :|                :    |||
Db         333 MKGLLC-DPNRLPPYLRWCVLSTPEIQKCGDM---------------AVAFRRQRLKPE 375

Qy         214 MKILSEVTP 222
               :: :|  :|
Db         376 IQCVSAKSP 384


RESULT 15
US-09-285-040-3
; Sequence 3, Application US/09285040
; Patent No. 6455494
;  GENERAL INFORMATION:
;    APPLICANT:  Jefferies, Wilfred A.
;    APPLICANT:  McGeer, Patrick L.
;    APPLICANT:  Rothenberger, Sylvia
;    APPLICANT:  Food, Michael R.
;    APPLICANT:  Yamada, Tatsuo
;    APPLICANT:  Kennard, Malcolm
;    TITLE OF INVENTION:  Use of p97 and Iron Binding Proteins
;    TITLE OF INVENTION:  as Diagnostic and Therapeutic Agents
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bereskin & Parr
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/285,040
;      FILING DATE:  01-APR-1999
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gravelle, Micheline
;      REGISTRATION NUMBER:  40,261
;      REFERENCE/DOCKET NUMBER:  7685-032
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  416-364-7311
;      TELEFAX:  416-361-1398
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  719 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-285-040-3

  Query Match             6.4%;  Score 77;  DB 2;  Length 719;
  Best Local Similarity   23.7%;  Pred. No. 19;
  Matches   59;  Conservative   25;  Mismatches   87;  Indels   78;  Gaps   15;

Qy         17 RGDFCGGTERAIDQASFTTSMEW----------DTQVVKGSSPLGPAGLGAEEPAAGP 64
              |||  |  |  |::      ::           |  || |: |       |   |
Db        171 RGDSSG--EGVCDKSPLERYYDYSGAFRCLAEGAGDVAFVKHSTVL--------ENTDGK 220

Qy         65 QLPSWLQP------------ERCAVFQCAQCH--AVLADSVHLAWDLSRSLGAVVFSRV 109
               |||| |              |  | :  |||   | | :| :  |     | ::| |:
Db        221 TLPSWGQALLSQDFELLCRDGSRADVTEWRQCHLARVPAHAVVVRADTD---GGLIF-RL 276

Qy        110 TNNVVLEAPFLVGIEGS----LKGSTY---NLLFCGSCG--IPVGFHLYST---HAALAA 157
              |    |   |   |||          |   :|||  |    :|:    |      |  | |
Db        277 LN----EGQRLFSHEGSSFQMFSSEAYGQKDLLFKDSTSELVPIATQTYEAWLGHEYLHA 332

Qy        158 LRGHFCLSSDKMVCYL----LKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLKSL 213
              ::|  |   :::  ||    | |  |    :|                :     |||
Db        333 MKGLLC-DPNRLPPYLRWCVLSTPEIQKCGDM---------------AVAFRRQRLKPE 375

Qy        214 MKILSEVTP 222
              :: :|  :|
Db        376 IQCVSAKSP 384


Search completed: November 18, 2005, 20:35:56
Job time : 50 secs
```

This Page Blank (use

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 18, 2005, 20:35:17 ; Search time 166 Seconds
(without alignments)
576.403 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1198	100.0	229	3	US-09-942-052-728	Sequence 728, App
2	1198	100.0	229	3	US-09-942-052-729	Sequence 729, App
3	1198	100.0	229	3	US-09-942-052-730	Sequence 730, App
4	1198	100.0	229	3	US-09-942-052-731	Sequence 731, App
5	138.5	11.6	164	3	US-09-942-052-707	Sequence 707, App
6	138	11.5	29	3	US-09-942-052-704	Sequence 704, App
7	113	9.4	233	3	US-09-890-688-82	Sequence 82, Appl
8	113	9.4	233	4	US-10-408-765A-969	Sequence 969, App
9	95.5	8.0	708	4	US-10-437-963-195427	Sequence 195427,
10	92	7.7	361	4	US-10-437-963-154548	Sequence 154548,
11	92	7.7	695	4	US-10-369-493-18389	Sequence 18389, A
12	89	7.4	664	3	US-09-780-525-2	Sequence 2, Appli
13	88.5	7.4	435	4	US-10-437-963-186569	Sequence 186569,
14	87.5	7.3	848	4	US-10-437-963-155606	Sequence 155606,
15	86.5	7.2	669	4	US-10-282-122A-51802	Sequence 51802, A
16	85	7.1	337	4	US-10-156-761-10954	Sequence 10954, A
17	84.5	7.1	485	4	US-10-363-829-358	Sequence 358, App
18	84	7.0	295	4	US-10-424-599-223506	Sequence 223506,
19	84	7.0	1902	4	US-10-437-963-139559	Sequence 139559,
20	83.5	7.0	888	4	US-10-437-963-147897	Sequence 147897,
21	83	6.9	663	4	US-10-282-122A-51673	Sequence 51673, A
22	82.5	6.9	124	4	US-10-425-115-216789	Sequence 216789,
23	82.5	6.9	278	4	US-10-425-114-42310	Sequence 42310, A
24	82.5	6.9	325	4	US-10-663-896-2	Sequence 2, Appli
25	82.5	6.9	493	5	US-10-213-974-36	Sequence 36, Appl
26	82.5	6.9	502	4	US-10-424-599-226730	Sequence 226730,
27	82.5	6.9	1043	4	US-10-369-493-9793	Sequence 9793, Ap
28	81.5	6.8	278	4	US-10-425-114-66220	Sequence 66220, A
29	81.5	6.8	457	4	US-10-195-518-6	Sequence 6, Appli
30	81.5	6.8	701	5	US-10-450-763-31748	Sequence 31748, A
31	81.5	6.8	1969	4	US-10-437-963-115342	Sequence 115342,
32	81.5	6.8	3067	5	US-10-631-467-1618	Sequence 1618, Ap
33	81	6.8	229	4	US-10-425-115-281194	Sequence 281194,
34	81	6.8	859	4	US-10-210-281-60	Sequence 60, Appl
35	80.5	6.7	282	5	US-10-450-763-51058	Sequence 51058, A
36	80.5	6.7	1711	5	US-10-029-345A-38	Sequence 38, Appl
37	80	6.7	305	4	US-10-183-116-59	Sequence 59, Appl
38	80	6.7	305	5	US-10-957-135-59	Sequence 59, Appl
39	80	6.7	305	6	US-11-083-611-59	Sequence 59, Appl
40	80	6.7	544	4	US-10-424-599-158084	Sequence 158084,
41	79.5	6.6	212	3	US-09-925-299-968	Sequence 968, App
42	79.5	6.6	212	3	US-09-925-299-968	Sequence 968, App
43	79	6.6	196	5	US-10-450-763-57058	Sequence 57058, A
44	79	6.6	230	5	US-10-450-763-44924	Sequence 44924, A
45	79	6.6	271	5	US-10-450-763-44919	Sequence 44919, A

ALIGNMENTS

RESULT 1
US-09-942-052-728
; Sequence 728, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Eid, Pia M.
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: 60/228,432
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 744
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 728
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: 85P1B3/OIP5
; OTHER INFORMATION: clone A protein
US-09-942-052-728

Query Match 100.0%; Score 1198; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 9.8e-114;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

```
RESULT 2
US-09-942-052-729
; Sequence 729, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 729
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3/OIP5
;   OTHER INFORMATION: protein sequence
US-09-942-052-729

  Query Match             100.0%;  Score 1198;  DB 3;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-114;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229


RESULT 3
US-09-942-052-730
; Sequence 730, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 730
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3 protein
;   OTHER INFORMATION: sequence
US-09-942-052-730

  Query Match             100.0%;  Score 1198;  DB 3;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-114;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229


RESULT 4
US-09-942-052-731
; Sequence 731, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 731
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-942-052-731

  Query Match             100.0%;  Score 1198;  DB 3;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 9.8e-114;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

```
RESULT 5
US-09-942-052-707
; Sequence 707, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 707
;   LENGTH: 164
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: Splice variant 1
;   OTHER INFORMATION: open reading frame 3 peptide sequence
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (44)
;   OTHER INFORMATION: Unknown amino acid or stop
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (50)..(51)
;   OTHER INFORMATION: Unknown amino acid or stop
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (67)
;   OTHER INFORMATION: Unknown amino acid or stop
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (93)
;   OTHER INFORMATION: Unknown amino acid or stop
US-09-942-052-707

  Query Match             11.6%;  Score 138.5;  DB 3;  Length 164;
  Best Local Similarity   82.1%;  Pred. No. 1.2e-05;
  Matches   32;  Conservative    1;  Mismatches    1;  Indels    5;  Gaps    1;

Qy        162 FCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              | |||     |||||||||||||||||||||||||||:|
Db          2 FFLSS-----YLLKTKAIVNASEMDIQNVPLSEKIAEVK 35


RESULT 6
US-09-942-052-704
; Sequence 704, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 704
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3 peptide
;   OTHER INFORMATION: sequence
US-09-942-052-704

  Query Match             11.5%;  Score 138;  DB 3;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        172 YLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              |||||||||||||||||||||||||||||
Db          1 YLLKTKAIVNASEMDIQNVPLSEKIAELK 29


RESULT 7
US-09-890-688-82
; Sequence 82, Application US/09890688
; Publication No. US20030144475A1
; GENERAL INFORMATION:
;  APPLICANT: Seishi KATO
;  APPLICANT:  Chikashi EGUCHI
;  APPLICANT:  Mihoro SAEKI
;  TITLE OF INVENTION: Human Proteins and cDNAs thereof
;  FILE REFERENCE: 2001-1102A/WMC/00653
;  CURRENT APPLICATION NUMBER: US/09/890,688
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: JP 11-346863
;  PRIOR FILING DATE: 1999-12-06
;  PRIOR APPLICATION NUMBER: JP 11-34684
;  PRIOR FILING DATE: 1999-12-06
;  PRIOR APPLICATION NUMBER: JP 2000-31062
;  PRIOR FILING DATE: 2000-02-08
;  PRIOR APPLICATION NUMBER: JP 2000-34091
;  PRIOR FILING DATE: 2000-02-10
;  PRIOR APPLICATION NUMBER: JP 2000-34090
;  PRIOR FILING DATE: 2000-02-10
;  PRIOR APPLICATION NUMBER: JP 2000-35829
;  PRIOR FILING DATE: 2000-02-14
;  PRIOR APPLICATION NUMBER: JP 2000-35899
;  PRIOR FILING DATE: 2000-02-14
;  PRIOR APPLICATION NUMBER: JP 2000-71161
;  PRIOR FILING DATE: 2000-03-14
;  PRIOR APPLICATION NUMBER: JP 2000-160851
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 160
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
;   LENGTH: 233
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-890-688-82

  Query Match             9.4%;  Score 113;  DB 3;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0075;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:      ::    |: || ::    :         |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209
```

```
Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221
```

RESULT 8
US-10-408-765A-969
; Sequence 969, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 969
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-969

Query Match 9.4%; Score 113; DB 4; Length 233;
Best Local Similarity 27.1%; Pred. No. 0.0075;
Matches 52; Conservative 30; Mismatches 88; Indels 22; Gaps 8;

```
Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV---ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:    ::   |: || ::   :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221
```

RESULT 9
US-10-437-963-195427
; Sequence 195427, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 195427
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_91377C.1.pep
US-10-437-963-195427

Query Match 8.0%; Score 95.5; DB 4; Length 708;
Best Local Similarity 24.5%; Pred. No. 2.1;
Matches 71; Conservative 32; Mismatches 86; Indels 101; Gaps 17;

```
Qy          6 LRHRSRCATPPRGDF-------------------------CGGTERAIDQASFTTSMEW 39
              |||| : |||   |                .          |||  |      |::
Db        225 LRHRRKAATPSASDASPPPPRRQSIVTIGEKEARAKAARAKSGGTSSA-----SPTTVST 279

Qy         40 DTQVVKGSSPLGPAGLGAEEPAAGPQLP----SW--LQPERCAVF---------QCAQCH 84
              |   | ||  : |:|    :|| || ||    :|  || |   :          : ::
Db        280 DVVPVVGSQEVTPSG-PISDPAGGPSLPEAVLTWEELQVEMGRLLKAGARGIGREISEAR 338

Qy         85 AVLADSVHLAWDLSRSLG----------AVVFSRVTNNVVLE------APFLVGIEGSLK 128
              | | :  ||  | | |            :|      :  ||          | | |||:
Db        339 AETAAANALAERLVRELAEARERDLTKMRELVAGNERQSKGLEDRMSELGDNLSEIRGSLR 398

Qy        129 GSTYNLL--FCGSCG----IPVG---FHLYSTHAALAALRG-----HFCLSSDKM----- 169
              ||  |     ||    |||    | | |: | ||   |     |    :::
Db        399 -VTYTGLHQLAKECGVKSTIPVNLDEFSLTSSLAELATAMGEIPSKHTSRIAEETSNGIY 457

Qy        170 --VCYLLKTKAIVNAS--EMDIQNV----------PLSEKIAELKEKIV 204
                ||:|   | |  |  |:|:: :          : |::::| | ::
Db        458 TGVCHVL---ACVRLSRPELDLREILDQGAASDTRKEVMEEVSDLGESVL 504
```

RESULT 10
US-10-437-963-154548
; Sequence 154548, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 154548
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54398C.1.pep
US-10-437-963-154548

Query Match 7.7%; Score 92; DB 4; Length 361;
Best Local Similarity 25.3%; Pred. No. 1.9;
Matches 61; Conservative 28; Mismatches 92; Indels 60; Gaps 14;

```
Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPA 61
              |  ||  |          |  |||  |    : |     |  || ||    |:|  | :|
Db         49 APTPLPFRRAVRAKAAQDGSGGTSSASPAVAST-----DVVVVPGSREATPSG-PASDPV 102

Qy         62 AGPQLP----SW--LQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVL 115
              ||    |    ||  || |   : :      |:   :  | |   :     | : ||
Db        103 AGRGSPAAVLSWEELQVEMGRLLEAGA--RVIGREIAEARGLEHRM-----SELGNN--- 152

Qy        116 EAPFLVGIEGSLKGSTYNLL--FCGSCGI-------PVGFHLYSTHAALAAL-----RGH 161
                |  | |||:  ||  |    | |||        |  | | |: | |||       |
Db        153 ----LSEIRGSLR-VTYTGLHQLAGKCGIKSTIPANPDEFSLTSSLAELAAAMEEIPSKH 207
```

```
Qy        162 FCLSSDKM-------VCYLLKTKAIVNASEMDIQNV-----------PLSEKIAELKEKI 203
                  ::|        |::|    : .:  |:|:: :           : |:: :| : :
Db        208 AARIGEEMSNRIYIGACHILACVRLAH-PELDLREILDQGEASDARKDVMEEVGDLGKSV 266

Qy        204 V 204
              :
Db        267 L 267


RESULT 11
US-10-369-493-18389
; Sequence 18389, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;  APPLICANT: Cao, Yongwei
;  APPLICANT:  Hinkle, Gregory J.
;  APPLICANT:  Slater, Steven C.
;  APPLICANT:  Goldman, Barry S.
;  APPLICANT:  Chen, Xianfeng
;  TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;  TITLE OF INVENTION:  PLANTS WITH IMPROVED PROPERTIES
;  FILE REFERENCE: 38-10(52052)B
;  CURRENT APPLICATION NUMBER: US/10/369,493
;  CURRENT FILING DATE:  2003-02-28
;  PRIOR APPLICATION NUMBER: US 60/360,039
;  PRIOR FILING DATE: 2002-02-21
;  NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18389
;   LENGTH: 695
;   TYPE: PRT
;   ORGANISM: Lactococcus lactis
US-10-369-493-18389

  Query Match             7.7%;  Score 92;  DB 4;  Length 695;
  Best Local Similarity   25.3%;  Pred. No. 4.7;
  Matches   38;  Conservative   28;  Mismatches   50;  Indels   34;  Gaps    7;

Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
              || ::|           ||  | :| ||     ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI------------QNVPLSE-K 195
              |         |    :: :|:: |||: | :   |:             ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
              : :|   :|  : ::  ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241


RESULT 12
US-09-780-525-2
; Sequence 2, Application US/09780525
; Patent No. US20020004223A1
; GENERAL INFORMATION:
;  APPLICANT: Bin-Bing Zhou
;  APPLICANT:  Yuan Zhu
;  APPLICANT:  Priya Chaturvedi
;  APPLICANT:  Mark R. Hurle
;  APPLICANT:  Xiaotong Li
;  TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
;  FILE REFERENCE: GP-70668-C1
;  CURRENT APPLICATION NUMBER: US/09/780,525
;  CURRENT FILING DATE:  2001-02-09
;  PRIOR APPLICATION NUMBER: 09/456,876
;  PRIOR FILING DATE: 1999-12-08
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 664
;   TYPE: PRT
;   ORGANISM: HOMO SAPIENS
US-09-780-525-2

  Query Match             7.4%;  Score 89;  DB 3;  Length 664;
  Best Local Similarity   23.0%;  Pred. No. 8.9;
  Matches   52;  Conservative   26;  Mismatches   74;  Indels   74;  Gaps   11;

Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|: |      | ||:::     ::||  |
Db        434 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 490

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||           |||  |   |    || |  :|
Db        491 PMPDRRAERE QDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 534

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||       | ||   |:  |      : :| |: |:|               | :
Db        535 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 578

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
              :|  || | ||       | ::||    ::    :    ||:| ||
Db        579 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 624


RESULT 13
US-10-437-963-186569
; Sequence 186569, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 186569
;   LENGTH: 435
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_83357C.1.pep
US-10-437-963-186569

  Query Match             7.4%;  Score 88.5;  DB 4;  Length 435;
  Best Local Similarity   22.9%;  Pred. No. 5.6;
  Matches   50;  Conservative   27;  Mismatches   82;  Indels   59;  Gaps   12;

Qy          2 AAQPL--RH--RSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGA 57
              | :||  ||    |||          ||  |  :: |           :  || |  |
Db         41 APEPLSCRHGRHLRCAAVD-----GGAGRETERPSPPAPQ-------REESPSGSLGAAL 88

Qy         58 EEPAAGPQLPSWLQPERCAVFQCAQCHA--VLADSVHLAWDLSRSLGAVVFSRVTNNVVL 115
              |:|:  | :|  | |        | :|:|  : :: |    ||    :       : :| :
Db         89 EDPSPQPGVPLILLP------LCCRCYAKHICSEYVVRTTDLVNHI-------LNSNAIS 135

Qy        116 EAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLK 175
              | ||           |     | ||       | :  |            | ::|   || |:
Db        136 EGPF----------SMRKARFLGSAS---AFSVKQTEWP--------CATTDSKKCY-LQ 173

Qy        176 TKAIVNASEMDIQN------VPLSEKIAELKEKIVLTH 207
                  :   :| : |:       |  | : : ||  :  |
Db        174 NGSFGGITEDEEQSSLYNFLYPSKELLPDDKEMSIFDH 211


RESULT 14
US-10-437-963-155606
```

; Sequence 155606, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 155606
; LENGTH: 848
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(848)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_55353C.1.pep
US-10-437-963-155606

Query Match 7.3%; Score 87.5; DB 4; Length 848;
Best Local Similarity 22.1%; Pred. No. 18;
Matches 65; Conservative 23; Mismatches 109; Indels 97; Gaps 13;

```
Qy          16 PRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP---------------- 60
               |||    || : : | :         |  :|||  ||| | |   |
Db          85 PRGFRLDGTGQGVRQTASCHLTRDLTDFLKGSGCLGPRGRGKRNPTPPPLGGNQEEGRHL 144

Qy          61 AAGPQLPSWLQPERCAVFQCAQCH-----------------AVLADSVHLAWDLSR---- 99
                 || :|   :|         | |                :  | |    |:|
Db         145 TLGPDVPEGARPRGRLCLPQASGHDTPGPISPTSLVLNKRTRIEIDEVVATMDISEANEG 204

Qy         100 --SLGAVV-FSRVTNNV-----VLEAPFLVGIEGSLK-GSTYNLLFCG------------ 138
                 | |:|:  ||               |   ||    | ||   |||
Db         205 YVSCGSVIEMSRQMKAARAGVRAQSXPACXGIPTVGKVGSV--LLFTARMGEGLCHMFTT 262

Qy         139 -----SCGIPV-----------------GFHLYSTH---AALAALRGHFCLSSDKMVCYL 173
                    ||: :                 |:     |      | | || |    | : |
Db         263 GSKAMECGVKMCLVGWPWLAWDELGPRSGYQFGLNHRRPLVKAVLDGHLCEEEDAVNCLP 322

Qy         174 LKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKP 227
                 | : :|||          |  :  |    :::| :|:|| |: | | |
Db         323 KLRKTVGSASE----------AKEAVKPAVKQEKKIRS-VKVLSLVS-DLSLP 363
```

RESULT 15
US-10-282-122A-51802
; Sequence 51802, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51802
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51802

Query Match 7.2%; Score 86.5; DB 4; Length 669;
Best Local Similarity 27.4%; Pred. No. 16;
Matches 37; Conservative 22; Mismatches 45; Indels 31; Gaps 8;

```
Qy          96 DLSR---SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNL------LFCGSCGIPVGF 146
               |: |    :|: || | :|:|:   | ::|:    :| |  :       :|||  :  |
Db         366 DIGRKKVKIGSRVFVRRSNDVI---PEIMGVTEETEGETNEIEAPTICPYCGSEIVKEGV 422

Qy         147 HLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP-LSEKIAE-LKEKIV 204
               ||              ||   : : |     |:||: :  :  |:   ||| || | ||
Db         423 HL------------FC--ENTLSCKPQMVKSIVHFASREAMNIEGFSEKTAEQLPEK-- 465

Qy         205 LTHNRLKSLMKILSE 219
               |    :  | :|  |
Db         466 LNIKSISDLYRITKE 480
```

Search completed: November 18, 2005, 20:49:37
Job time : 168 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 18, 2005, 20:36:06 ; Search time 7 Seconds
(without alignments)
36.960 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 8323 seqs, 1129788 residues

Total number of hits satisfying chosen parameters: 8323

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	68	5.7	1255	7	US-11-022-562-235	Sequence 235, App
2	67.5	5.6	1992	7	US-11-013-759-3	Sequence 3, Appli
3	67.5	5.6	1992	7	US-11-013-759-13	Sequence 13, Appl
4	67.5	5.6	2047	7	US-11-013-759-4	Sequence 4, Appli
5	67.5	5.6	2047	7	US-11-013-759-7	Sequence 7, Appli
6	67	5.6	381	1	US-10-689-742-172	Sequence 172, App
7	66	5.5	480	1	US-10-336-263A-8	Sequence 8, Appli
8	65	5.4	680	1	US-10-467-962B-101	Sequence 101, App
9	62.5	5.2	747	1	US-10-131-826A-212	Sequence 212, App
10	62	5.2	592	1	US-10-467-962B-95	Sequence 95, Appl
11	60.5	5.1	669	1	US-10-997-201A-30	Sequence 30, Appl
12	60	5.0	614	7	US-11-074-176-80	Sequence 80, Appl
13	58.5	4.9	639	7	US-11-074-176-222	Sequence 222, App
14	58	4.8	250	1	US-10-512-184-27	Sequence 27, Appl
15	58	4.8	421	1	US-10-131-826A-302	Sequence 302, App
16	57.5	4.8	389	1	US-10-979-821-2	Sequence 2, Appli
17	56.5	4.7	129	1	US-10-816-768-89	Sequence 89, Appl
18	56.5	4.7	409	1	US-10-627-633-4	Sequence 4, Appli
19	55.5	4.6	413	1	US-10-957-569-9	Sequence 9, Appli
20	55.5	4.6	732	1	US-10-518-599-22	Sequence 22, Appl
21	55	4.6	101	1	US-10-816-768-51	Sequence 51, Appl
22	55	4.6	414	1	US-10-131-826A-418	Sequence 418, App
23	55	4.6	442	7	US-11-074-176-282	Sequence 282, App
24	54.5	4.5	1013	1	US-10-131-826A-38	Sequence 38, Appl
25	54	4.5	231	1	US-10-957-569-21	Sequence 21, Appl
26	54	4.5	751	7	US-11-012-762-26	Sequence 26, Appl
27	53.5	4.5	101	1	US-10-816-768-49	Sequence 49, Appl
28	53.5	4.5	379	1	US-10-131-826A-308	Sequence 308, App
29	53.5	4.5	415	1	US-10-627-633-2	Sequence 2, Appli
30	53	4.4	352	1	US-10-627-633-6	Sequence 6, Appli
31	53	4.4	388	7	US-11-082-389-220	Sequence 220, App
32	53	4.4	433	1	US-10-652-893-4	Sequence 4, Appli
33	53	4.4	547	7	US-11-082-389-86	Sequence 86, Appl
34	53	4.4	660	1	US-10-131-826A-480	Sequence 480, App
35	53	4.4	746	7	US-11-074-176-314	Sequence 314, App
36	53	4.4	749	7	US-11-074-176-54	Sequence 54, Appl
37	52.5	4.4	102	1	US-10-816-768-52	Sequence 52, Appl
38	52.5	4.4	106	1	US-10-816-768-62	Sequence 62, Appl
39	52.5	4.4	106	1	US-10-816-768-63	Sequence 63, Appl
40	52.5	4.4	271	7	US-11-082-389-208	Sequence 208, App
41	52.5	4.4	358	1	US-10-689-742-170	Sequence 170, App
42	52.5	4.4	509	1	US-10-131-826A-108	Sequence 108, App
43	52.5	4.4	585	1	US-10-967-457-18	Sequence 18, Appl
44	52.5	4.4	585	1	US-10-939-890-500	Sequence 500, App
45	52.5	4.4	690	1	US-10-939-890-501	Sequence 501, App

ALIGNMENTS

RESULT 1
US-11-022-562-235
; Sequence 235, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 235
; LENGTH: 1255
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
US-11-022-562-235

Query Match 5.7%; Score 68; DB 7; Length 1255;
Best Local Similarity 21.8%; Pred. No. 8;
Matches 36; Conservative 24; Mismatdhes 69; Indels 36; Gaps 7;

```
Qy          29 DQASFTTSM-EWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQC---- 83
               | : || |: :  |  :   ||    |:  |           |   ::|   |
Db         554 DVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTN------ASSEVAVLYQDVNCTDVS 607

Qy          84 HAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIP 143
                 |: || :  ||         ::|   |    :|  |:| |     ::|       | ||
Db         608 TAIHADQLTPAWR--------IYSTGNNVFQTQAGCLIGAEH--VDTSY------ECDIP 651

Qy         144 VGFHLYSTHAALAALRG---------HFCLSSDKMVCYLLKTKAI 179
               :|  : :::  :: ||           | :|  : |   | ||
Db         652 IGAGICASYHTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAI 696
```

RESULT 2
US-11-013-759-3
; Sequence 3, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Sasaki, Ken

```
;  APPLICANT:  Yang, Yan Ping
;  APPLICANT:  Klein, Michel H.
;  TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
;  TITLE OF INVENTION:  PROTEIN OF MORAXELLA
;  FILE REFERENCE: 1038-921MIS:jb
;  CURRENT APPLICATION NUMBER: US/11/013,759
;  CURRENT FILING DATE:  2004-12-16
;  PRIOR APPLICATION NUMBER: US/09/361,619
;  PRIOR FILING DATE: 1999-07-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1992
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-11-013-759-3

  Query Match             5.6%;  Score 67.5;  DB 7;  Length 1992;
  Best Local Similarity   22.0%;  Pred. No. 18;
  Matches   40;  Conservative   20;  Mismatches   47;  Indels   75;  Gaps    7;

Qy         46 GSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVV 105
              | | |   ||  : |    |:                    | || |
Db        964 GKSTLNDGGLSIKNPTGSEQI-----------------QVGADGVK------------ 992

Qy        106 FSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS 165
              |::| || |: |    ||:|: : :   : | |: |                      |
Db        993 FAKVNNNGVVGA----GIDGTTRITRDEIGFTGTNG----------------------S 1025

Qy        166 SDKMVCYLLKTKAIVNASEMDIQNVPLSE-----------KIAELKEKIVLTHNRLKSL 213
                ||   :|  :|  :||    | |:   |            || :|| ::    |:: |
Db       1026 LDKSKPHL--SKDGINAGGKKITNIQSGEIAQNSHDAVTGGKIYDLKTEL---ENKISST 1080

Qy        214 MK 215
               |
Db       1081 AK 1082


RESULT 3
US-11-013-759-13
; Sequence 13, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  APPLICANT:  Sasaki, Ken
;  APPLICANT:  Yang, Yan Ping
;  APPLICANT:  Klein, Michel H.
;  TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
;  TITLE OF INVENTION:  PROTEIN OF MORAXELLA
;  FILE REFERENCE: 1038-921MIS:jb
;  CURRENT APPLICATION NUMBER: US/11/013,759
;  CURRENT FILING DATE:  2004-12-16
;  PRIOR APPLICATION NUMBER: US/09/361,619
;  PRIOR FILING DATE: 1999-07-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 1992
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-11-013-759-13

  Query Match             5.6%;  Score 67.5;  DB 7;  Length 1992;
  Best Local Similarity   22.0%;  Pred. No. 18;
  Matches   40;  Conservative   20;  Mismatches   47;  Indels   75;  Gaps    7;

Qy         46 GSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVV 105
              | | |   ||  : |    |:                    | || |
Db        964 GKSTLNDGGLSIKNPTGSEQI-----------------QVGADGVK------------ 992

Qy        106 FSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS 165
              |::| || |: |    ||:|: : :   : | |: |                      |
Db        993 FAKVNNNGVVGA----GIDGTTRITRDEIGFTGTNG----------------------S 1025

Qy        166 SDKMVCYLLKTKAIVNASEMDIQNVPLSE-----------KIAELKEKIVLTHNRLKSL 213
                ||   :|  :|  :||    | |:   |            || :|| ::    |:: |
Db       1026 LDKSKPHL--SKDGINAGGKKITNIQSGEIAQNSHDAVTGGKIYDLKTEL---ENKISST 1080

Qy        214 MK 215
               |
Db       1081 AK 1082


RESULT 4
US-11-013-759-4
; Sequence 4, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  APPLICANT:  Sasaki, Ken
;  APPLICANT:  Yang, Yan Ping
;  APPLICANT:  Klein, Michel H.
;  TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
;  TITLE OF INVENTION:  PROTEIN OF MORAXELLA
;  FILE REFERENCE: 1038-921MIS:jb
;  CURRENT APPLICATION NUMBER: US/11/013,759
;  CURRENT FILING DATE:  2004-12-16
;  PRIOR APPLICATION NUMBER: US/09/361,619
;  PRIOR FILING DATE: 1999-07-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 2047
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-11-013-759-4

  Query Match             5.6%;  Score 67.5;  DB 7;  Length 2047;
  Best Local Similarity   22.0%;  Pred. No. 18;
  Matches   40;  Conservative   20;  Mismatches   47;  Indels   75;  Gaps    7;

Qy         46 GSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVV 105
              | | |   ||  : |    |:                    | || |
Db       1019 GKSTLNDGGLSIKNPTGSEQI-----------------QVGADGVK------------ 1047

Qy        106 FSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS 165
              |::| || |: |    ||:|: : :   : | |: |                      |
Db       1048 FAKVNNNGVVGA----GIDGTTRITRDEIGFTGTNG----------------------S 1080

Qy        166 SDKMVCYLLKTKAIVNASEMDIQNVPLSE-----------KIAELKEKIVLTHNRLKSL 213
                ||   :|  :|  :||    | |:   |            || :|| ::    |:: |
Db       1081 LDKSKPHL--SKDGINAGGKKITNIQSGEIAQNSHDAVTGGKIYDLKTEL---ENKISST 1135

Qy        214 MK 215
               |
Db       1136 AK 1137


RESULT 5
US-11-013-759-7
; Sequence 7, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
;  APPLICANT: Loosmore, Sheena M.
;  APPLICANT:  Sasaki, Ken
;  APPLICANT:  Yang, Yan Ping
;  APPLICANT:  Klein, Michel H.
;  TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
;  TITLE OF INVENTION:  PROTEIN OF MORAXELLA
;  FILE REFERENCE: 1038-921MIS:jb
;  CURRENT APPLICATION NUMBER: US/11/013,759
;  CURRENT FILING DATE:  2004-12-16
;  PRIOR APPLICATION NUMBER: US/09/361,619
;  PRIOR FILING DATE: 1999-07-27
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 7
;   LENGTH: 2047
;   TYPE: PRT
;   ORGANISM: Moraxella catarrhalis
US-11-013-759-7

  Query Match             5.6%;  Score 67.5;  DB 7;  Length 2047;
  Best Local Similarity   22.0%;  Pred. No. 18;
  Matches   40;  Conservative   20;  Mismatches   47;  Indels   75;  Gaps    7;

Qy         46 GSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVV 105
Db       1019 GKSTLNDGGLSIKNPTGSEQI-----------------QVGADGVK------------ 1047

Qy        106 FSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLS 165
Db       1048 FAKVNNNGVVGA----GIDGTTRITRDEIGFTGTNG----------------------S 1080

Qy        166 SDKMVCYLLKTKAIVNASEMDIQNVPLSE-----------KIAELKEKIVLTHNRLKSL 213
Db       1081 LDKSKPHL--SKDGINAGGKKITNIQSGEIAQNSHDAVTGGKIYDLKTEL---ENKISST 1135

Qy        214 MK 215
Db       1136 AK 1137


RESULT 6
US-10-689-742-172
; Sequence 172, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
;  APPLICANT: Jacobs, Kenneth
;  APPLICANT:  McCoy, John  M
;  APPLICANT:  LaVallie, Edward R
;  APPLICANT:  Racie, Lisa A
;  APPLICANT:  Evans, Cheryl
;  APPLICANT:  Merberg, David
;  APPLICANT:  Treacy, Maurice
;  APPLICANT:  Spaulding, Vikki
;  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
;  FILE REFERENCE: 00766.000091.10
;  CURRENT APPLICATION NUMBER: US/10/689,742
;  CURRENT FILING DATE:  2003-10-22
;  PRIOR APPLICATION NUMBER: 09/746,783
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 172
;   LENGTH: 381
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-689-742-172

  Query Match             5.6%;  Score 67;  DB 1;  Length 381;
  Best Local Similarity   21.7%;  Pred. No. 1.8;
  Matches   23;  Conservative   19;  Mismatches   42;  Indels   22;  Gaps    4;

Qy         78 FQCAQCH-------AVLADSVHL-----AWDLSRSLGAVVFSRVTNNVVLEAPFLVGIEG 125
Db       202 FKCPQCNNRKEFPQEMLRMGIHIPDRDAAWELEPGAFSDLYQRYQHCDAPICPYEQGRDS 261

Qy        126 SLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVC 171
Db       262 FEDEGRWCLILCATCG------SHGTHRDCSSLR----FNSKKWEC 297


RESULT 7
US-10-336-263A-8
; Sequence 8, Application US/10336263A
; Publication No. US20050251882A1
; GENERAL INFORMATION:
;  APPLICANT: D'Ordine, Robert L
;  APPLICANT:  Dotson, Stanton B
;  APPLICANT:  Duff, Stephen M
;  APPLICANT:  Sisson, Pamela J
;  TITLE OF INVENTION: SUCROSE PHOSPHATE SYNTHASE NUCLEIC ACID MOLECULES FROM CYANOBACTE
;  TITLE OF INVENTION:  USES THEREFOR
;  FILE REFERENCE: 38-21 (52176) A
;  CURRENT APPLICATION NUMBER: US/10/336,263A
;  CURRENT FILING DATE:  2003-01-03
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
;   LENGTH: 480
;   TYPE: PRT
;   ORGANISM: Nostoc punctiforme
;   FEATURE:
;   NAME/KEY: PEPTIDE
;   LOCATION: (1)..(480)
;   OTHER INFORMATION:
US-10-336-263A-8

  Query Match             5.5%;  Score 66;  DB 1;  Length 480;
  Best Local Similarity   28.9%;  Pred. No. 3.3;
  Matches   22;  Conservative   14;  Mismatches   30;  Indels   10;  Gaps    3;

Qy        149 YSTHAALAALRGHF-CLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
Db        407 FSTQGVAARVHELYQSLTLDTFIQEIIKTKKLTPDLERQIQNL--------LKSK-VLKS 457

Qy        208 NRLKSLMKILSEVTPD 223
Db        458 NEIKALEKLIDSFSND 473


RESULT 8
US-10-467-962B-101
; Sequence 101, Application US/10467962B
; Publication No. US20050246784A1
; GENERAL INFORMATION:
;  APPLICANT: Plesch, Gunnar
;  APPLICANT:  Blau, Astrid
;  APPLICANT:  Daeschner, Klaus
;  APPLICANT:  Klein, Mathieu
;  TITLE OF INVENTION: Identification of Herbicidally Active Substances
;  FILE REFERENCE: 2000_857
;  CURRENT APPLICATION NUMBER: US/10/467,962B
;  CURRENT FILING DATE:  2003-08-14
;  PRIOR APPLICATION NUMBER: PCT/EP02/01466
;  PRIOR FILING DATE: 2002-02-13
;  NUMBER OF SEQ ID NOS: 109
;  SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 101
;   LENGTH: 680
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-467-962B-101

  Query Match             5.4%;  Score 65;  DB 1;  Length 680;
  Best Local Similarity   24.6%;  Pred. No. 7;
  Matches   43;  Conservative   25;  Mismatches   81;  Indels   26;  Gaps    8;

Qy         73 ERCAVFQCAQCHAVLADSV--HLAW-DLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKG 129
Db        477 ERRGCLPCATTYNIMINSYGDQERWDDMKRLLGKMKSQGILPNVVTHTT-LVDVYG--KS 533

Qy        130 STYN-----LLFCGSCGIPVGFHLYSTHAALAALRG--------HFCLSSDKMVCYLLKT 176
Db        534 GRFNDAIECLEEMKSVGLKPSSTMYNALINAYAQRGLSEQAVNAFRVMTSDGLKPSLLAL 593

Qy        177 KAIVNASEMDIQNVPLSEKIAELKEK----IVLTHNRLKSLMKILSEVTPDQSKP 227
Db        594 NSLINAFGEDRRDAEAFAVLQYMKENGVKPDVVTYT---TLMKALIRVDKFQKVP 645
```

```
RESULT 9
US-10-131-826A-212
; Sequence 212, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C128
;  CURRENT APPLICATION NUMBER: US/10/131,826A
;  CURRENT FILING DATE:  2002-04-24
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE: 1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE: 1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE: 1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE: 1997-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 212
;   LENGTH: 747
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-131-826A-212

  Query Match             5.2%;  Score 62.5;  DB 1;  Length 747;
  Best Local Similarity   24.3%;  Pred. No. 15;
  Matches   26;  Conservative   18;  Mismatches   38;  Indels   25;  Gaps    5;

Qy        137 CGSC-----GIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP 191
              | :|      :||    :   | |  |: ||  |:      :|   |  :|:  |:|||
Db        509 CSACHNERLDVPV----WDVEATLNFLKAHFSPSN-----IILDFPAAGSAARRDVQNVA 559

Qy        192 LSEKIA----ELKEKIVLTHNRLKSLMKILSEVT-------PDQSKP 227
                : ::|    ||: :          :||  :  |       |: |:|
Db        560 AAPELAMGALELESRNSTLDPGKPEMMKSPTNTTPHVPAEGPEASRP 606


RESULT 10
US-10-467-962B-95
; Sequence 95, Application US/10467962B
; Publication No. US20050246784A1
; GENERAL INFORMATION:
;  APPLICANT: Plesch, Gunnar
;  APPLICANT:  Blau, Astrid
;  APPLICANT:  Daeschner, Klaus
;  APPLICANT:  Klein, Mathieu
;  TITLE OF INVENTION: Identification of Herbicidally Active Substances
;  FILE REFERENCE: 2000_857
;  CURRENT APPLICATION NUMBER: US/10/467,962B
;  CURRENT FILING DATE:  2003-08-14
;  PRIOR APPLICATION NUMBER: PCT/EP02/01466
;  PRIOR FILING DATE: 2002-02-13
;  NUMBER OF SEQ ID NOS: 109
;  SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 95
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-10-467-962B-95

  Query Match             5.2%;  Score 62;  DB 1;  Length 592;
  Best Local Similarity   23.0%;  Pred. No. 12;
  Matches   32;  Conservative   22;  Mismatches   37;  Indels   48;  Gaps    7;

Qy        103 AVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAA-------- 154
              ||||:| :    |::     :: :: |    : || | |  :| :: |
Db        325 AVVFNRSSRFARLQS-----VKCTIAGKNAYVRFCCSTGDAMGMNMVSKGVQNVLEYLTD 379

Qy        155 ------LAALRGHFCLSSDK-------------MVCYLLKTKAIVNASEMDIQNVPLSE 194
                   :  : |:||  |||                :||     :|::     :| |  |
Db        380 DFPDMDVIGISGNFC--SDKKPAAVNWIEGRGKSVVC-----EAVIRG---EIVNKVLKT 429

Qy        195 KIAELKEKIVLTHNRLKSL 213
               :| | |      | ||:|
Db        430 SVAALVEL-----NMLKNL 443


RESULT 11
US-10-997-201A-30
; Sequence 30, Application US/10997201A
; Publication No. US20050249739A1
; GENERAL INFORMATION:
;  APPLICANT: Marasco, Wayne
;  APPLICANT:  Sui, Jianhua
;  TITLE OF INVENTION: Antibodies Against SARS-COV and Methods of Use Thereof
;  FILE REFERENCE: 20363-026
;  CURRENT APPLICATION NUMBER: US/10/997,201A
;  CURRENT FILING DATE:  2004-11-24
;  PRIOR APPLICATION NUMBER: 60/524,840
;  PRIOR FILING DATE: 2003-11-25
;  NUMBER OF SEQ ID NOS: 40
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
;   LENGTH: 669
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-997-201A-30

  Query Match             5.1%;  Score 60.5;  DB 1;  Length 669;
  Best Local Similarity   22.1%;  Pred. No. 21;
  Matches   30;  Conservative   22;  Mismatches   57;  Indels   27;  Gaps    6;

Qy         29 DQASFTTSM-EWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQC---- 83
              | : || |: :  |  :   ||    |:    |                |   ::|   |
Db        550 DVSDFTDSVRDPKTSEILDISPCAFGGVSVITPGTN------ASSEVAVLYQDVNCTDVS 603

Qy         84 HAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIP 143
               |: || :  ||         ::|   |    :|  |:| |     ::|       | ||
Db        604 TAIHADQLTPAWR--------IYSTGNNVFQTQAGCLIGAEH--VDTSY------ECDIP 647

Qy        144 VGFHLYSTHAALAALR 159
              :|  : :::  :: ||
Db        648 IGAGICASYHTVSLLR 663
```

RESULT 12
US-11-074-176-80
; Sequence 80, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 614
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-80

Query Match 5.0%; Score 60; DB 7; Length 614;
Best Local Similarity 19.9%; Pred. No. 21;
Matches 46; Conservative 33; Mismatches 100; Indels 52; Gaps 9;

```
Qy          25 ERAIDQASPTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCH 84
               : |:  |  | : : |  :: | |   ||   | :  || :   : |:       |
Db         292 DNALKDAGLTVN-DIDKVILNGGSTRIPAVQKAVKEWAGKEPDHSINPDEAVALGAA--- 347

Qy          85 AVLADSVHLAWDLSRSLGAVVFSRVTNNVVLE-APFLVGIEGSLKGSTYNLLFCGSCGIP 143
                          | |:   | : |:|:  |  :|||    |  :  |   :  ||
Db         348 ---------------IQGGVISGDVKDIVLLDVTPLSLGIE--TMGGVFTKLIDRNTTIP 390

Qy         144 VG-FHLYSTHA------ALAALRGHFCLSSDK--------------------MVCYLLK 175
                 ::|| |        :  |:|   :::|                      | : :
Db         391 TSKSQIFSTAADNQPAVDVHVLQGERPMAADDKTLGRFELTDIPPAPRGVPQIQVTFDID 450

Qy         176 TKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSK 226
                  ||| |  |:     :  :|| :|    |:   :| : |   |   :  |
Db         451 KNGIVNVSAKDM-GTGKEQKIT-IKSSSGLSDEEIKRMQKDAEEHAEEDKK 499
```

RESULT 13
US-11-074-176-222
; Sequence 222, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 222
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-222

Query Match 4.9%; Score 58.5; DB 7; Length 639;
Best Local Similarity 22.5%; Pred. No. 33;
Matches 56; Conservative 31; Mismatches 71; Indels 91; Gaps 17;

```
Qy          25 ERAI--DQASFTTSMEWD------------TQVVKGSSPLGPAGLGAEEPAAGPQLPSW 69
               || |     :|: :: |:             |  | | :         || |     | |
Db         160 ERGIYTSLGAFSGAIGWNGLAIIVVPLVTGVTYAVTGKN---------EEGA-----PGW 205

Qy          70 LQPER--------CAVFQC---AQCHAVLADSV----------------HLAWDLSRSL 101
               |            ||:  |    : | :: :|                    : |    ||
Db         206 LAFAAVISALAIICAIIVCFGTKEKHNLIRNSAKQKTTLRQVFSAIFHNDQILWP---SL 262

Qy         102 GAVVFS---RVTNNV-------VLEAP---FLVGIEGSLKGSTYNLLFCGSCGIPVGFHL 148
                 :|||    :|| |       |:  |   ::||:  :: |      || |   |:  :
Db         263 AYLVFSLANTITNGVLFYLYKFVIGKPGEFWVVGLIATIIG------FCVSPMFPI-LNK 315

Qy         149 YSTHAALAALRGHFCLSSDKMVC-YLLKTKAIVNASEMDIQNVPLSEKIAELKEKIVLT- 206
               |     |  : |  |     |:| |||      |   ||:  :  :   |:|   : ||
Db         316 YIPRKWL-FIAGQTC-----MICAYLLFIFCRDNVVLMDLGLILFNINFAQLVTVLTLTD 369

Qy         207 ---HNRLKS 212
                  : :|||
Db         370 AIEYGQLKS 378
```

RESULT 14
US-10-512-184-27
; Sequence 27, Application US/10512184
; Publication No. US20050244901A1
; GENERAL INFORMATION:
; APPLICANT: Fraunhofer Gesellschaft zur F"rderung der angewandten Forschung e.V.
; TITLE OF INVENTION: Antibodies, recombinant antibodies, recombinant
; TITLE OF INVENTION: antibody fragments and fusions mediated plant disease
; TITLE OF INVENTION: resistance against fungi
; FILE REFERENCE: 3581.01US01
; CURRENT APPLICATION NUMBER: US/10/512,184
; CURRENT FILING DATE: 2004-10-22
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: scFv SGB3 with
; OTHER INFORMATION: specificity against Fusarium ssp.; originates from
; OTHER INFORMATION: Gallus gallus.
US-10-512-184-27

Query Match 4.8%; Score 58; DB 1; Length 250;
Best Local Similarity 24.5%; Pred. No. 9.5;
Matches 27; Conservative 15; Mismatches 40; Indels 28; Gaps 5;

```
Qy          22 GGTERAIDQASPTTSMEWDTQVVKGSSPL--------GPAGL-----GAEEPAAGPQLPS 68
               ||| : |  :  |    |  |   ||:|:         |: :     |::  : |    :
Db         156 GGTVK-ITCSGSTAHYSWHQQKSPGSAPVTLISFNNQRPSDIPSRFSGSKSGSTGTLTIT 214

Qy          69 WLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAP 118
                 :: |  ||: |              ||  ||: | :|    |   ||  |
Db         215 GVRAEDEAVYYCG-----------GWD--RSITAGLFGAGTTLTVLGQP 250
```

RESULT 15
US-10-131-826A-302
; Sequence 302, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey

```
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C128
;  CURRENT APPLICATION NUMBER: US/10/131,826A
;  CURRENT FILING DATE:  2002-04-24
;  PRIOR APPLICATION NUMBER: 60/049911
;  PRIOR FILING DATE: 1997-06-18
;  PRIOR APPLICATION NUMBER: 60/056974
;  PRIOR FILING DATE: 1997-08-26
;  PRIOR APPLICATION NUMBER: 60/059113
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059115
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059117
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059122
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059184
;  PRIOR FILING DATE: 1997-09-17
;  PRIOR APPLICATION NUMBER: 60/059263
;  PRIOR FILING DATE: 1997-09-18
;  PRIOR APPLICATION NUMBER: 60/059352
;  PRIOR FILING DATE: 1997-09-19
;  PRIOR APPLICATION NUMBER: 60/059588
;  PRIOR FILING DATE: 1997-09-19
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 302
;   LENGTH: 421
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-131-826A-302

  Query Match             4.8%;  Score 58;  DB 1;  Length 421;
  Best Local Similarity   32.6%;  Pred. No. 20;
  Matches   28;  Conservative    9;  Mismatches   27;  Indels   22;  Gaps    6;

Qy        141 GIPVGFHLYSTHAALAALRGHFCLSS--DKMVCYLLKTKAIVNASEMDIQNVPLSEKIAE 198
              |:|  || |:    | : :   || ::    | |   | | | |   | || : |     |
Db         52 GVP--FHQYT----LGCVSECFCPTNFPSSMYCDNRKLKTIPNI-PMHIQQLYLQFNEIE 104

Qy        199 ------------LKEKIVLTHNRLKS 212
                         ||| | |:||::||
Db        105 AVTANSFINATHLKE-INLSHNKIKS 129


Search completed: November 18, 2005, 20:49:50
Job time : 8 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 18, 2005, 20:27:55 ; Search time 40 Seconds
(without alignments)
550.841 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	92	7.7	695	2	C86731	copper-potassium t
2	90.5	7.6	597	2	D71293	probable phosphori
3	89	7.4	306	2	T46399	hypothetical prote
4	88	7.3	250	2	A43623	kanamycin kinase (
5	86.5	7.2	669	2	A97229	NAD-dependent DNA
6	85	7.1	2194	1	JQ1977	glutamate synthase
7	84.5	7.1	361	2	AE2074	ferrichrome bindin
8	83	6.9	663	2	D97047	DNA ligase (NAD de
9	82.5	6.9	143	2	B69099	formate hydrogenly
10	82.5	6.9	493	2	A85433	sugar transporter
11	82	6.8	967	2	F87678	DNA polymerase I [
12	81.5	6.8	329	2	D90404	transport protein,
13	81.5	6.8	457	2	H85767	glucuronide permea
14	81.5	6.8	457	2	C90919	glucuronide permea
15	81.5	6.8	457	2	B64918	glucuronide permea
16	81.5	6.8	640	2	T41977	hypothetical prote
17	81.5	6.8	662	2	T44221	probable capsid pr
18	81	6.8	1418	2	S40764	hypothetical prote
19	80.5	6.7	766	2	B85440	receptor kinase-li
20	80.5	6.7	1711	1	A55148	protein-tyrosine-p
21	80	6.7	469	2	S55167	IME2-dependent sig
22	79	6.6	386	2	H90789	probable aminometh
23	79	6.6	386	2	C85650	probable aminometh
24	79	6.6	437	2	C86823	GTP-binding protei
25	78.5	6.6	188	2	F83816	late competence op
26	78.5	6.6	578	2	F86484	probable hydroxyme
27	78.5	6.6	697	1	QQBEK3	UL104 protein - hu
28	78	6.5	351	2	H82098	UDP-3-O-3-hydroxym
29	77.5	6.5	662	2	T44036	hypothetical prote
30	77.5	6.5	1013	2	T31211	trwC protein homol
31	77.5	6.5	1643	2	T14274	versican precursor
32	77	6.4	381	2	AD1113	hypothetical prote
33	77	6.4	721	2	F87611	TonB-dependent rec
34	77	6.4	738	1	TFHUM	melanotransferrin
35	76.5	6.4	307	2	AG2017	glycerol-3-phospha
36	76.5	6.4	387	2	D69392	probable acyl-CoA
37	76.5	6.4	388	2	S57526	cellulase - Fibrob
38	76.5	6.4	764	2	H98143	cbbBc protein (U60
39	76.5	6.4	764	2	AD3144	formate dehydrogen
40	76.5	6.4	1554	2	T06370	probable DNA (cyto
41	76	6.3	367	2	AH0936	glycerol dehydroge
42	75.5	6.3	310	2	A55053	endothelial monocy
43	75.5	6.3	415	2	T46716	hypothetical prote
44	75.5	6.3	700	2	D70951	probable UvrD - My
45	75.5	6.3	1534	2	S59604	DNA (cytosine-5-)-

ALIGNMENTS

RESULT 1
C86731
copper-potassium transporting ATPase B copB [imported] - Lactococcus lactis subsp. lact
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C;Accession: C86731
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86731
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-695 <STO>
A;Cross-references: UNIPROT:Q9CH87; UNIPARC:UPI00000C6919; GB:AE005176; PID:g12723778;
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: copB
C;Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding

Query Match 7.7%; Score 92; DB 2; Length 695;
Best Local Similarity 25.3%; Pred. No. 2;
Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

```
Qy      101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
            || ::|              ||  | :| ||    ::   : || | :  || :| : :| |
Db      101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy      161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI-------------QNVPLSE-K 195
            |          |    :: :|:: |||: | :   |:              ::| ||| |
Db      154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy      196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
                : :|   :|  : :: :   ||||   |:|
Db      214 VGDL--LLVKENEKIPADGLILSEALVDES 241
```

RESULT 2
D71293
probable phosphoribosylglycinamide formyltransferase - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004
C;Accession: D71293
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwi
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71293
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

```
A;Residues: 1-597 <COL>
A;Cross-references: UNIPROT:O83693; UNIPARC:UPI00000D3290; GB:AE001243; GB:AE000520; NID
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0695

  Query Match             7.6%;  Score 90.5;  DB 2;  Length 597;
  Best Local Similarity   28.0%;  Pred. No. 2.2;
  Matches   51;  Conservative   19;  Mismatches   47;  Indels   65;  Gaps   12;

Qy         12 CATPPRGDFCGGTERAID----QASFT-------------TSMEWDTQVVKGSSPLGPA 53
              || |  |     |: | |    :| ||             |  |||       | | |
Db         90 CALP--GHRLEATKNATDKTRMRACFTRARLRCPRFTFLEPDSFAWDT-------PPGHA 140

Qy         54 GLGAEEPAAGPQLPSWLQP-ERCAVFQC--AQCHAVLADSVHLAWDLSRSLGAVVFSRVT 110
               | :   :||   |  ::| :      |  |||   | ::  :|   |||
Db        141 RLCSHLHSAGLSFPLVVKPTDNMGARGCTLAQCKDTLINACAVARQFSRS---------- 190

Qy        111 NNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMV 170
              |::|  |:|| | ||:|    |:| |:        || |  |||           |: :
Db        191 GRVIIE-EFIVGREFSLEG----LIFDGT--------LYVT--ALA----------DRHI 225

Qy        171 CY 172
              |:
Db        226 CF 227


RESULT 3
T46399
hypothetical protein DKFZp434N2420.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46399
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46399
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-306 <AAA>
A;Cross-references: UNIPROT:Q96EP1; UNIPARC:UPI000006D779; EMBL:AL137561
A;Experimental source: adult testis; clone DKFZp434N2420
C;Genetics:
A;Note: DKFZp434N2420.1

  Query Match             7.4%;  Score 89;  DB 2;  Length 306;
  Best Local Similarity   23.0%;  Pred. No. 1.3;
  Matches   52;  Conservative   26;  Mismatches   74;  Indels   74;  Gaps   11;

Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |      |  :|:  |     | ||:::     ::||  |
Db         76 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 132

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||             |||  |   |    || |  :|
Db        133 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 176

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|               | :
Db        177 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 220

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
               :|  || | ||       | ::||   ::    :    ||:| ||
Db        221 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 266


RESULT 4
A43623
kanamycin kinase (EC 2.7.1.95) - Campylobacter jejuni
C;Species: Campylobacter jejuni
C;Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 31-Dec-2004
C;Accession: A43623
R;Tenover, F.C.; Gilbert, T.; O'Hara, P.
Plasmid 22, 52-58, 1989
A;Title: Nucleotide sequence of a novel kanamycin resistance gene, aphA-7, from Campyloba
A;Reference number: A43623; MUID:89387451; PMID:2550983
A;Accession: A43623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <TEN>
A;Cross-references: UNIPROT:P14508; UNIPARC:UPI000012DEBC; GB:M29953; GB:J03316; NID:g144
C;Superfamily: aminoglycoside 3'-phosphotransferase (kanamycin kinase)
C;Keywords: phosphotransferase

  Query Match             7.3%;  Score 88;  DB 2;  Length 250;
  Best Local Similarity   28.7%;  Pred. No. 1.3;
  Matches   33;  Conservative   18;  Mismatches   46;  Indels   18;  Gaps    5;

Qy        105 VFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG-HFC 163
              :||: | :|  ||  :: :   ||              |  |   :| :  :: ||| |
Db         40 IFSKTTYSVKREAEMMMWLSDKLKVPDV---------IEYGVREHSEYLIMSELRGKHID 90

Qy        164 LSSDKMVCYLLKTKAIVNA----SEMDIQNVPLSEKI-AELKEKIVLTHNRLKSL 213
                  | : |:   : :|||      :||:| | | ||   |||    |  ||:  :
Db         91 CFIDHPIKYI---ECLVNALHQLQAIDIRNCPFSSKIDVRLKELKYLLDNRIADI 142


RESULT 5
A97229
NAD-dependent DNA ligase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 31-Dec-2004
C;Accession: A97229
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clos
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97229
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-669 <KUR>
A;Cross-references: UNIPROT:Q97FQ5; UNIPARC:UPI00000CA5D7; GB:AE001437; PIDN:AAK80620.1;
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2673
C;Superfamily: DNA ligase (NAD), LigA type

  Query Match             7.2%;  Score 86.5;  DB 2;  Length 669;
  Best Local Similarity   27.4%;  Pred. No. 6.2;
  Matches   37;  Conservative   22;  Mismatches   45;  Indels   31;  Gaps    8;

Qy         96 DLSR---SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNL------LFCGSCGIPVGF 146
              |: |    :|: || | :|:|:    | ::|:     :| |  :        :|||  :  |
Db        366 DIGRKKVKIGSRVFVRRSNDVI---PEIMGVTEETEGETNEIEAPTICPYCGSEIVKEGV 422

Qy        147 HLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP-LSEKIAE-LKEKIV 204
              ||              ||   : : |    |:||: :   :   |:   ||| || | ||
Db        423 HL-------------FC--ENTLSCKPQMVKSIVHFASREAMNIEGFSEKTAEQLFEK-- 465

Qy        205 LTHNRLKSLMKILSE 219
              |     :  | :|  |
Db        466 LNIKSISDLYRITKE 480


RESULT 6
JQ1977
glutamate synthase (NADH2) (EC 1.4.1.14) precursor [validated] - alfalfa
C;Species: Medicago sativa (alfalfa)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-Oct-2004
C;Accession: JQ1977; PQ0551
R;Gregerson, R.G.; Miller, S.S.; Twary, S.N.; Gantt, J.S.; Vance, C.P.
Plant Cell 5, 215-226, 1993
A;Title: Molecular characterization of NADH-dependent glutamate synthase from alfalfa no
A;Reference number: JQ1977; MUID:93200806; PMID:8453303
A;Accession: JQ1977
```

A;Molecule type: mRNA
A;Residues: 1-2194 <GRE>
A;Cross-references: UNIPROT:Q03460; UNIPARC:UPI000012B7FA; GB:L01660; NID:g166411; PIDN:
A;Accession: PQ0551
A;Molecule type: protein
A;Residues: 102-114 <GR2>
A;Cross-references: UNIPARC:UPI0000172061
C;Comment: This enzyme catalyzes the reductive transfer of the amido group of glutamine
C;Superfamily: glutamate synthase (NADH/NADPH), eukaryotic type
C;Keywords: 3Fe-4S; chloroplast; iron-sulfur protein; metalloprotein; NAD; oxidoreductas
F;1-101/Domain: propeptide #status predicted <PRO>
F;102-2194/Product: glutamate synthase (NADH) #status experimental <MAT>
F;102/Active site: Cys #status predicted
F;1246,1252,1257/Binding site: 3Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 7.1%; Score 85; DB 1; Length 2194;
Best Local Similarity 24.7%; Pred. No. 39;
Matches 58; Conservative 22; Mismatches 79; Indels 76; Gaps 12;

```
Qy          4 QPLRHRSRCATPPRGDFCGGTERAIDQAS----------FTTSMEWDTQVVKGSSPLGPA 53
              :||   ||   | |          || | :          | : |   :: :|: |
Db       1037 EPLADGSR--NPKRS--------AIKQVASGRFGVSSYYLTNADELQIKMAQGAKP---- 1082

Qy         54 GLGAEEP-------------AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR 99
              | | | |            ||  | |   |    ::       : |    |  ||
Db       1083 GEGGELPGHKVIGDIAITRNSTAGVGLIS--PPPHHDIYS-------IEDLAQLIHDLKN 1133

Qy        100 SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCG-------------SCGIPVG 145
              :  |   :|::  :| ||    |   | :||   ::|  |                | |:|
Db       1134 ANPA---ARISVKLVSEAGVGVIASGVVKGHAEHVLISGHDGGTGASRWTGIKSAGLPWE 1190

Qy        146 FHLYSTHAALAA--LRGHFCLSSDKMVCYLLKT------KAIVNASEMDIQNVPL 192
              |  ||  | |  |||   | :|      |||       |:: | |      ||
Db       1191 LGLAETHQTLVANDLRGRTTLQTDGQ----LKTGRDVAIAALLGAEEYGFSTAPL 1241
```

RESULT 7
AE2074
ferrichrome binding protein of ABC transporter all2147 [imported] - Nostoc sp. (strain P
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AE2074
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE2074
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-361 <KUR>
A;Cross-references: UNIPROT:Q8YV34; UNIPARC:UPI00000CE32E; GB:BA000019; PIDN:BAB73846.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2147
C;Superfamily: ferrichrome-iron transport protein fecB

Query Match 7.1%; Score 84.5; DB 2; Length 361;
Best Local Similarity 26.1%; Pred. No. 4.3;
Matches 36; Conservative 17; Mismatches 36; Indels 49; Gaps 7;

```
Qy         97 LSRSLGAVVF----SRVT---NNVVLEAPFLVGIEGSLKGSTYNLLFCGSC-----GIPV 144
              :| :|| |        ||     |::|::   :|::          :::|  |     |||
Db         81 ISHALGKVKIPLKPQRVVVLEENIILDSVLALGVK------PVGVMYCQDCEENFRGIP- 133

Qy        145 GFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELKEKIV 204
                                 ||      ||    :|     :| | |  |||  ||  ::
Db        134 --------------------SD-----LLADVPVVG----NIGNQPSLEKILSLKPDLI 163

Qy        205 LTHNRLKSLMKILSEVTP 222
              |    |||  |||| : |
Db        164 LGLTWLKSSYKILSSIAP 181
```

RESULT 8
D97047
DNA ligase (NAD dependent), LigA [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 31-Dec-2004
C;Accession: D97047
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97047
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-663 <KUR>
A;Cross-references: UNIPROT:Q97JS8; UNIPARC:UPI00000CA10F; GB:AE001437; PIDN:AAK79167.1;
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1195
C;Superfamily: DNA ligase (NAD), LigA type

Query Match 6.9%; Score 83; DB 2; Length 663;
Best Local Similarity 28.6%; Pred. No. 13;
Matches 36; Conservative 18; Mismatches 44; Indels 28; Gaps 7;

```
Qy        102 GAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNL------LFCGSCGIPVGFHLYSTHAAL 155
              || || | :|:|:   | ::|:    :| |  :        :|||  :  | ||
Db        368 GARVFLRRSNDVI---PEIMGVTEETEGETKEIEAPTICPYCGSEIVKEGVHL------- 417

Qy        156 AALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP-LSEKIAE-LKEKIVLTHNRLKSL 213
                    ||   : : |      |:||: :    |:   ||| || | ||  |    :  |
Db        418 ------FC--ENTLSCKPQMVKSIVHFASRKAMNIEGFSEKTAEQLFEK--LNIKSISDL 467

Qy        214 MKILSE 219
                :|  |
Db        468 YRITKE 473
```

RESULT 9
B69099
formate hydrogenlyase, iron-sulfur subunit 2 - Methanobacterium thermoautotrophicum (str
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: B69099
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: B69099
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-143 <MTH>
A;Cross-references: UNIPROT:O27769; UNIPARC:UPI0000066781; GB:AE000929; GB:AE000666; NID
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1736
C;Superfamily: nrfC protein; ferredoxin 2[4Fe-4S] homology
F;62-118/Domain: ferredoxin 2[4Fe-4S] homology <FER2>

Query Match 6.9%; Score 82.5; DB 2; Length 143;
Best Local Similarity 23.0%; Pred. No. 2.1;
Matches 35; Conservative 26; Mismatches 66; Indels 25; Gaps 6;

```
Qy         72 PERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVF------SRVTNNVVLEAPFLVGIEG 125
              || |   :| :|  :   :      |     | ||          |     :    :| ::|
Db          8 PELCD--ECMKCERICPKNAIRVID-----GVPVFCMHCSPERAPCLNICPEDAIVEVDG 60

Qy        126 SLKGSTYNLLFCGSC--GIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNAS 183
              ::       : || |    |||    : :    |::   |:  || :| ::  |  :: |
```

```
Db         61 AVVILEDRCIGCGLCRDACPVG--AITLNERGVAVKCDLCIDRDKPLCVMVCPKGALSES 118

Qy        184 EMDIQNVPLSEKIAELKEKIVLTHNRLKSLMK 215
              |:        :|  ::||     |||:|||
Db        119 SEDM--------MAAKRDKIAGELKRLKNLMK 142
```

RESULT 10
A85433
sugar transporter like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A85433
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: A85433
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-493 <STO>
A;Cross-references: UNIPROT:O23213; UNIPARC:UPI000009F22F; GB:NC_001268; NID:g7270615; P
C;Genetics:
A;Gene: AT4g36670
A;Map position: 4
C;Superfamily: glucose transport protein

  Query Match 6.9%; Score 82.5; DB 2; Length 493;
  Best Local Similarity 24.9%; Pred. No. 10;
  Matches 53; Conservative 32; Mismatches 59; Indels 69; Gaps 13;

```
Qy         79 QCAQCHAVLADSVHLAWDLSRSLGAVVF---SRVTNNVVLEAPFLVGI------EGSLK- 128
              |||   :::  |:   :|     ||:||      ||:| :|   | ||       |||
Db         19 QCAIVASIV--SIIFGYDTGVMSGAMVFIEEDLKTNDVQIEV--LTGILNLCALVGSLLA 74

Qy        129 ----------------------------GSTYNLLFCGSC--GIPVGFHL-----YSTH 152
                                         |  | :|  | |  |: ||| |     ||
Db         75 GRTSDIIGRRYTIVLASILFMLGSILMGWGPNYPVLLSGRCTAGLGVGFALMVAPVYSAE 134

Qy        153 AALAALRG------HFCLSSDKMVCYLLK---TKAIVNAS---EMDIQNVP---LSEKIA 197
              | |: ||      | |:|   :: |::    :|  ::      : |  ||    |:  |
Db        135 IATASHRGLLASLPHLCISIGILLGYIVNYFFSKLPMHIGWRLMLGIAAVPSLVLAFGIL 194

Qy        198 ELKE--KIVLTHNRLKSLMKILSEV--TPDQSK 226
              :: |  : ::   |||   :||  |  :|::::
Db        195 KMPESPRWLIMQGRLKEGKEILELVSNSPEEAE 227
```

RESULT 11
F87678
DNA polymerase I [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: F87678
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-967 <STO>
A;Cross-references: UNIPROT:Q9A2U2; UNIPARC:UPI00000C7AB1; GB:AE005673; NID:g13425184; P
C;Genetics:
A;Gene: CC3464
C;Superfamily: DNA-directed DNA polymerase I

  Query Match 6.8%; Score 82; DB 2; Length 967;
  Best Local Similarity 24.1%; Pred. No. 26;
  Matches 51; Conservative 27; Mismatches 72; Indels 62; Gaps 11;

```
Qy         12 CATPPRGDFCGGTERAIDQ---ASFTTSMEWDT---QVVKGSSPLGPAGLGAEEPAAGPQ 65
              | ||         | |  |:   |:|   ||: :   :|  ||:   |  |  : ||| |:
Db        271 CDTPLPQPLDALTVREPDKEALAAFLEQMEFRSLARRVGDGSAAATPGTL--DRPAAPPK 328

Qy         66 LP----SWL-----------QPERC--AVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSR 108
              |    |::           :| :   | : | :  | |   |  | |     | | |
Db        329 APVVSVSYMGAAARAAAHPVEPVKIDHAAYACVRDLATLKAWVAKATD----KGLVAFDT 384

Qy        109 VTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDK 168
              |:              :|  :|      | ||:          :||   |  |
Db        385 ETD-------------ALSSAT-----AGLCGV----------SLAIAPGEACYIP-- 412

Qy        169 MVCYLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              : :  |   :   :  ||: :||:: || ||
Db        413 -ISHCEKADGLAFEAPADIEQIPLADVIATLK 443
```

RESULT 12
D90404
transport protein, probable [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: D90404
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-W
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: D90404
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-329 <KUR>
A;Cross-references: UNIPROT:Q97W97; UNIPARC:UPI00000646FA; GB:AE006641; NID:g13815639; PI
C;Genetics:
A;Gene: SSO2338

  Query Match 6.8%; Score 81.5; DB 2; Length 329;
  Best Local Similarity 22.2%; Pred. No. 7.4;
  Matches 40; Conservative 30; Mismatches 47; Indels 63; Gaps 7;

```
Qy         87 LADSVHLA-WD---------LSRSLGAVVFSRVTNNVV--------------------LE 116
              |::|:||| |:          | | :|: ::    |:|:                    :
Db         33 LSESMHLAYWEVFAIVALPFLGRIIGSFIYQVFKNSVISYCFPFLGFLVILQNFLGALIF 92

Qy        117 APFLVG----------IEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSS 166
              ||||           :| ::|     |:   : | |:|: :
Db         93 VRFLVGVIFGLLTSYAVESAVKSGRNVLVGFTTAGWPIGWVI------------------ 134

Qy        167 DKMVCY-LLKTKAIVNASEMDIQNVPLSE---KIAELKEKIVLTHNRLKSLMKILSEVTP 222
              | | |||   ::| | : |  : | |   |    : || ::  || |::  :| :||
Db        135 -SYVAYVLLKNWNVINISGILIMLLALFELNGKEFGERSKISVSFPRLTSILIYVSALTP 193
```

RESULT 13
H85767
glucuronide permease [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: H85767
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85767
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <STO>
A;Cross-references: UNIPROT:Q8X673; UNIPARC:UPI00000D0BE5; GB:AE005174; NID:g12515601; P
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: uidB

C;Superfamily: melibiose carrier protein

Query Match 6.8%; Score 81.5; DB 2; Length 457;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

```
Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
               :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456
```

RESULT 14
C90919
glucuronide permease [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: C90919
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90919
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <HAY>
A;Cross-references: UNIPROT:Q8X673; UNIPARC:UPI00000D0BE5; GB:BA000007; PIDN:BAB35746.1;
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2323
C;Superfamily: melibiose carrier protein

Query Match 6.8%; Score 81.5; DB 2; Length 457;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

```
Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
               :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456
```

RESULT 15
B64918
glucuronide permease uidB - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: B64918
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64918
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-457 <BLAT>
A;Cross-references: UNIPROT:P30868; UNIPARC:UPI0000137B23; GB:AE000257; GB:U00096; NID:g
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: uidB; gusB
C;Superfamily: melibiose carrier protein
C;Keywords: carrier protein; transmembrane protein
F;37-53/Domain: transmembrane #status predicted <TM01>
F;81-97/Domain: transmembrane #status predicted <TM02>
F;152-168/Domain: transmembrane #status predicted <TM03>
F;184-200/Domain: transmembrane #status predicted <TM04>
F;231-247/Domain: transmembrane #status predicted <TM05>
F;263-279/Domain: transmembrane #status predicted <TM06>
F;310-326/Domain: transmembrane #status predicted <TM07>
F;408-424/Domain: transmembrane #status predicted <TM08>

Query Match 6.8%; Score 81.5; DB 2; Length 457;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

```
Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
               :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456
```

Search completed: November 18, 2005, 20:36:43
Job time : 43 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: November 18, 2005, 20:24:56 ; Search time 232 Seconds
(without alignments)
696.405 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1198	100.0	229	1	OIP5_HUMAN	O43482 homo sapien
2	122	10.2	218	2	Q9CXR6_MOUSE	Q9cxr6 mus musculu
3	118	9.8	462	2	Q4PAB5_USTMA	Q4pab5 ustilago ma
4	115.5	9.6	204	2	Q9CZJ6_MOUSE	Q9czj6 mus musculu
5	113	9.4	233	1	CU045_HUMAN	Q9nyp9 homo sapien
6	113	9.4	233	2	Q542Z0_HUMAN	Q542z0 homo sapien
7	110.5	9.2	232	1	CU045_PANTR	Q68ut5 pan troglod
8	109	9.1	207	2	Q5BLB1_BRARE	Q5blb1 brachydanio
9	107	8.9	532	2	Q54HI6_DICDI	Q54hi6 dictyosteli
10	105	8.8	155	1	YCZC_SCHPO	Q9p802 schizosacch
11	94	7.8	1556	2	Q4Q626_LEIMA	Q4q626 leishmania
12	92	7.7	361	2	Q8S611_ORYSA	Q8s611 oryza sativ
13	92	7.7	695	2	Q9CH87_LACLA	Q9ch87 lactococcus
14	91	7.6	628	2	Q69KE0_ORYSA	Q69ke0 oryza sativ
15	90.5	7.6	597	2	O83693_TREPA	O83693 treponema p
16	89.5	7.5	502	2	Q50PD2_ENTHI	Q50pd2 entamoeba h
17	89.5	7.5	661	2	Q50V43_ENTHI	Q50v43 entamoeba h
18	89.5	7.5	661	2	Q50WZ8_ENTHI	Q50wz8 entamoeba h
19	89	7.4	633	2	Q5M052_STRT1	Q5m052 streptococc
20	89	7.4	664	1	CHFR_HUMAN	Q96ep1 homo sapien
21	89	7.4	735	2	Q5M4R3_STRT2	Q5m4r3 streptococc
22	89	7.4	893	1	POL2_BAMMA	Q65329 barley mild
23	88.5	7.4	152	2	Q50XW7_ENTHI	Q50xw7 entamoeba h
24	88.5	7.4	638	2	Q75RZ6_CIOIN	Q75rz6 ciona intes
25	88	7.3	250	1	KKA7_CAMJE	P14508 campylobact
26	88	7.3	493	2	Q68VH2_9POTY	Q68vh2 barley mild
27	87.5	7.3	634	2	Q6CX82_KLULA	Q6cx82 kluyveromyc
28	87.5	7.3	696	2	Q4JTS4_CORJK	Q4jts4 corynebacte
29	87	7.3	664	1	CHFR_MOUSE	Q81013 mus musculu
30	86.5	7.2	669	2	Q97FQ5_CLOAB	Q97fq5 clostridium
31	86.5	7.2	2093	2	Q5K2K3_9RHAB	Q5k2k3 isfahan vir
32	86.5	7.2	5344	2	Q529N3_MAGGR	Q529n3 magnaporthe
33	86	7.2	145	2	Q839T4_ENTFA	Q839t4 enterococcu
34	86	7.2	327	2	Q4NE64_9MICC	Q4ne64 arthrobacte
35	86	7.2	331	2	Q6GXE5_COCCA	Q6gxe5 cochliobolu
36	86	7.2	893	2	Q68VH4_9POTY	Q68vh4 barley mild
37	86	7.2	3692	2	Q5J1Q7_9NOCA	Q5j1q7 nocardia un
38	85	7.1	337	1	SYW1_STRAW	Q82hu1 streptomyce
39	85	7.1	671	2	Q8I5K5_PLAF7	Q8i5k5 plasmodium
40	85	7.1	1018	2	Q6FKN6_CANGA	Q6fkn6 candida gla
41	85	7.1	2194	1	GLSN_MEDSA	Q03460 medicago sa
42	85	7.1	2194	2	Q40360_MEDSA	Q40360 medicago sa
43	84.5	7.1	361	2	Q8YV34_ANASP	Q8yv34 anabaena sp
44	84.5	7.1	376	2	Q5YZE1_NOCFA	Q5yze1 nocardia fa
45	84	7.0	247	2	Q64WS0_BACFR	Q64ws0 bacteroides

ALIGNMENTS

RESULT 1
OIP5_HUMAN
ID OIP5_HUMAN STANDARD; PRT; 229 AA.
AC O43482; Q96BX7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Opa-interacting protein 5.
GN Name=OIP5;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98125741; PubMed=9466265;
RA Williams J.M., Chen G.-C., Zhu L., Rest R.F.;
RT "Using the yeast two-hybrid system to identify human epithelial cell
RT proteins that bind gonococcal Opa proteins: intracellular gonococci
RT bind pyruvate kinase via their Opa proteins and require host pyruvate
RT for growth.";
RL Mol. Microbiol. 27:171-186(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBUNIT: Binds outer membrane protein OpaP from Neisseria
CC gonorrhoeae.
CC -!- INTERACTION:
CC P04049:RAF1; NbExp=3; IntAct=EBI-536879, EBI-365996;
CC ----------
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ----------------------------------------------------------------------
DR   EMBL; AF025441; AAC39561.1; ALT_INIT; mRNA.
DR   EMBL; BC015050; AAH15050.1; -; mRNA.
DR   IntAct; O43482; -.
DR   Ensembl; ENSG00000104147; Homo sapiens.
DR   HGNC; HGNC:20300; OIP5.
DR   H-InvDB; HIX0012150; -.
DR   MIM; 606020; -.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0007154; P:cell communication; NAS.
SQ   SEQUENCE   229 AA;  24691 MW;  0EBD4006193A3106 CRC64;

  Query Match           100.0%;  Score 1198;  DB 1;  Length 229;
  Best Local Similarity  100.0%;  Pred. No. 8.8e-105;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229


RESULT 2
Q9CXR6_MOUSE
ID   Q9CXR6_MOUSE PRELIMINARY;      PRT;   218 AA.
AC   Q9CXR6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched
DE   library, clone:3110025H23 product:hypothetical protein, full insert
DE   sequence.
GN   Name=2610039C10Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK014084; BAB29147.1; -; mRNA.
DR   Ensembl; ENSMUSG00000022978; Mus musculus.
DR   MGI; MGI:1913828; 2610039C10Rik.
KW   Hypothetical protein.
SQ   SEQUENCE   218 AA;  24447 MW;  55390E5BFDDDE8C2 CRC64;

  Query Match            10.2%;  Score 122;  DB 2;  Length 218;
  Best Local Similarity  23.3%;  Pred. No. 0.0049;
  Matches   48;  Conservative   30;  Mismatches  102;  Indels   26;  Gaps    6;

Qy         29 DQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLA 88
              | : :    :|    :  :  ||       || ||  : |         || ||:|    |
Db         15 DSSRYLRLQKWAN--MSSADALGLEKERPEEKAAAAENP--------LVFLCARCRRPLG 64

Qy         89 DSVHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFH 147
              || | |  |:     ::   |: || ::    :        |      |:|  | : :|:
Db         65 DS--LTWVASQEDTNCILLRSVSCNVSVDKEPKLSKCRDEDGCILEALYCTGCSLSLGYV 122
```

```
Qy        148 LYSTHAALAALRGHFCLSSDKMVCYLL---------KTKAIVNASEMDIQNVPLSEKIAE 198
Db        123 YRCTPKNLDYKRDLFCLSVETVESYTLGSSEKQIVSKDKELFNLE----SRVEIEKSIKQ 178

Qy        199 LKEKIVLTHNRLKSLMKILSEVTPDQ 224
Db        179 MEEVLTALQKKLREVESKLSLAQPGQ 204


RESULT 3
Q4PAB5_USTMA
ID   Q4PAB5_USTMA PRELIMINARY;      PRT;   462 AA.
AC   Q4PAB5;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=UM02948.1;
OS   Ustilago maydis 521.
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=237631;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=521;
RA   Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA   Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA   Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA   Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA   Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA   Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA   Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA   David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA   Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA   Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA   Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA   Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA   Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA   Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA   Jaffe D., Jones C., Kamal M., Kamat A., Kamvysselis M., Karlsson E.,
RA   Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA   Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA   Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA   Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA   Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA   Mccarthy M., Mcdonough S., Mcghee T., Meldrim J., Meneus L.,
RA   Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA   Mozes J., Mulrain L., Munson G., Naylor J., Newes C., Nguyen C.,
RA   Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA   Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA   O'neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piqani B.,
RA   Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA   Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA   Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA   Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA   Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA   Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA   Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA   Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA   Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA   Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA   Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA   Zimmer A., Zody M., Lander E.;
RT   "The genome sequence of Ustilago maydis.";
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AACP01000101; EAK84120.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   462 AA;  49283 MW;  EDDCFDF17757B509 CRC64;

  Query Match             9.8%;  Score 118;  DB 2;  Length 462;
  Best Local Similarity   25.9%;  Pred. No. 0.03;
  Matches   64;  Conservative   28;  Mismatches  109;  Indels   46;  Gaps   10;

Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQV-VKGSSPLGPAGLGAEEP 60
Db        109 AAQPLSSDDDDMAPPKG-----------GSLTTT---STRARGRGRGRGRPRGRGRGRG 153

Qy         61 AA-------GPQL--PSWLQPERCAVFQCAQCHAVLADSV-HLAWDLSRSLGAVVFSRVT 110
Db        154 ASTSTTVVHKPDLEESDSEEPPNPLVFQCRCCFRLLGDSLAFVATDI--DLGYVILSDVS 211

Qy        111 NNVVLEAPFLVGIE-GSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKM 169
Db        212 EIIQQDTTYETSTEPGKDIGSTFARLRCAGCNAAVGRNYRTTPRDLDDLRDCFSLEVDAI 271

Qy        170 VCYLLKTKAIVNASEMDIQ----------------NVPLSEKIAELKEKIVLTH-NRLKS 212
Db        272 YTYQLGSNYTRQKEEDDEDQDALVDGGASGKPARASRPTHETVQAEDTKVLITKMERTRA 331

Qy        213 LMKILSE 219
Db        332 LTIELSD 338


RESULT 4
Q9CZJ6_MOUSE
ID   Q9CZJ6_MOUSE PRELIMINARY;      PRT;   204 AA.
AC   Q9CZJ6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   Mus musculus 11 days embryo whole body cDNA, RIKEN full-length
DE   enriched library, clone:2700078I24 product:hypothetical protein, full
DE   insert sequence (2610039C10Rik protein).
GN   Name=2610039C10Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
```

```
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6; TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [8]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6; TISSUE=Eye;
RA   Director MGC Project;
RL   Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK012533; BAB28302.1; -; mRNA.
DR   EMBL; BC079900; AAH79900.1; -; mRNA.
DR   Ensembl; ENSMUSG00000022978; Mus musculus.
DR   MGI; MGI:1913828; 2610039C10Rik.
KW   Hypothetical protein.
SQ   SEQUENCE   204 AA;  22949 MW;  BBD667250FDF9B0F CRC64;

  Query Match             9.6%;  Score 115.5;  DB 2;  Length 204;
  Best Local Similarity   25.6%;  Pred. No. 0.018;
  Matches   50;  Conservative   29;  Mismatches   97;  Indels   19;  Gaps    7;

Qy         29 DQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLA 88
              | : :    :|    :  :  ||       || ||  : |          || ||:|   |
Db         15 DSSRYLRLQKWAN--MSSADALGLEKERPEEKAAAAENP--------LVFLCARCRRPLG 64

Qy         89 DSVHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFH 147
              ||  | |  |:     ::    |: || ::   :        |     |:|  | : :|:
Db         65 DS--LTWVASQEDTNCILLRSVSCNVSVDKEPKLSKCRDEDGCILEALYCTGCSLSLGYV 122

Qy        148 LYSTHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVL 205
              |   |   |  |||| : :  | |    | || : : :: |:   |   |::: |
Db        123 YRCTPKNLDYKRDLFCLSVEAVESYTLGSSEKQIV-SEDKELFNL---ESRVEIEKSIKQ 178

Qy        206 THNRLKSLMKILSEV 220
              | :| | | ||
Db        179 MEEVLTALQKKLREV 193


RESULT 5
CU045_HUMAN
ID   CU045_HUMAN    STANDARD;      PRT;   233 AA.
AC   Q9NYP9;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Protein C21orf45.
GN   Name=C21orf45;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   MEDLINE=20237674; PubMed=10773462; DOI=10.1016/S0378-1119(00)00089-5;
RA   Slavov D., Hattori M., Sakaki Y., Rosenthal A., Shimizu N.,
RA   Minoshima S., Kudoh J., Yaspo M.-L., Ramser J., Reinhardt R.,
```

```
RA   Reimer C., Clancy K., Rynditch A., Gardiner K.;
RT   "Criteria for gene identification and features of genome organization:
RT   analysis of 6.5 Mb of DNA sequence from human chromosome 21.";
RL   Gene 247:215-232(2000).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC   TISSUE=Blood;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: To S.pombe SpC970.12.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AF231921; AAF72945.1; -; mRNA.
DR   EMBL; BC042917; AAH42917.1; -; mRNA.
DR   Ensembl; ENSG00000159055; Homo sapiens.
DR   HGNC; HGNC:1286; C21orf45.
SQ   SEQUENCE   233 AA;  25863 MW;  A07522806C4B6221 CRC64;

  Query Match             9.4%;  Score 113;  DB 1;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.038;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 6
Q542Z0_HUMAN
ID   Q542Z0_HUMAN PRELIMINARY;      PRT;   233 AA.
AC   Q542Z0;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein FLJ90800 (FAPP1-associated protein 1).
GN   Name=FASP1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Otsuki T., Ota T., Nishikawa T., Hayashi K., Suzuki Y., Yamamoto J.,
RA   Wakamatsu A., Kimura K., Sakamoto K., Hatano N., Kawai Y., Ishii S.,
RA   Saito K., Kojima S., Sugiyama T., Ono T., Okano K., Yoshikawa Y.,
RA   Aotsuka S., Sasaki N., Hattori A., Okumura K., Nagai K., Sugano S.,
RA   Isogai T.;
RT   "Signal Sequence and Keyword Trap in silico for Selection of Full-
RT   Length Human cDNAs Encoding Secretion or Membrane Proteins from Oligo-
RT   Capped cDNA Libraries.";
RL   DNA Res. 12:117-126(2005).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=21061410; PubMed=11001876; DOI=10.1042/0264-6021:3510019;
RA   Dowler S.J., Currie R.A., Campbell D.G., Deak M., Kular G.,
RA   Downes C.P., Alessi D.R.;
RT   "Identification of pleckstrin-homology-domain-containing proteins with
RT   novel phosphoinositide-binding specificities.";
RL   Biochem. J. 351:19-31(2000).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RA   Dowler S.J.;
RT   "Identification of FAPP1 interacting proteins.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK075281; BAC11517.1; -; mRNA.
DR   EMBL; AF387845; AAK70498.1; -; mRNA.
SQ   SEQUENCE   233 AA;  25863 MW;  A07522806C4B6221 CRC64;

  Query Match             9.4%;  Score 113;  DB 2;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.038;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 7
CU045_PANTR
ID   CU045_PANTR    STANDARD;      PRT;   232 AA.
AC   Q68UT5;
DT   01-FEB-2005 (Rel. 46, Created)
DT   01-FEB-2005 (Rel. 46, Last sequence update)
DT   13-SEP-2005 (Rel. 48, Last annotation update)
DE   Protein C21orf45 homolog.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX   PubMed=15164055; DOI=10.1038/nature02564;
RA   Watanabe H., Fujiyama A., Hattori M., Taylor T.D., Toyoda A.,
RA   Kuroki Y., Noguchi H., BenKahla A., Lehrach H., Sudbrak R., Kube M.,
```

```
RA   Taenzer S., Galgoczy P., Platzer M., Scharfe M., Nordsiek G.,
RA   Bloecker H., Hellmann I., Khaitovich P., Paabo S., Reinhardt R.,
RA   Zheng H.-J., Zhang X.-L., Zhu G.-F., Wang B.-F., Fu G., Ren S.-X.,
RA   Zhao G.-P., Chen Z., Lee Y.-S., Cheong J.-E., Choi S.-H., Wu K.-M.,
RA   Liu T.-T., Hsiao K.-J., Tsai S.-F., Kim C.-G., Oota S., Kitano T.,
RA   Kohara Y., Saitou N., Park H.-S., Wang S.-Y., Yaspo M.-L., Sakaki Y.;
RT   "DNA sequence and comparative analysis of chimpanzee chromosome 22.";
RL   Nature 429:382-388(2004).
CC   -!- SIMILARITY: To S.pombe SpC970.12.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AL954205; CAH18576.1; -; Genomic_DNA.
SQ   SEQUENCE   232 AA;  25832 MW;  6549B8BE369245E7 CRC64;

  Query Match             9.2%;  Score 110.5;  DB 1;  Length 232;
  Best Local Similarity   27.1%;  Pred. No. 0.065;
  Matches   52;  Conservative   29;  Mismatches   88;  Indels   23;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||           ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERARL--EEAAAA--------EERPLVFLCSGCRRPLGDS 94

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:      ::   |: || ::    :         |      | |  | : :|:
Db         95 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 152

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  ||||  : :  |:|    | || : : :: |:   |   |::: :
Db        153 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 208

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        209 DVLKALQMKLWE 220


RESULT 8
Q5BLB1_BRARE
ID   Q5BLB1_BRARE PRELIMINARY;      PRT;   207 AA.
AC   Q5BLB1;
DT   10-MAY-2005 (TrEMBLrel. 30, Created)
DT   10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT   10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE   LOC553502 protein (Fragment).
GN   Name=LOC553502;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Singapore local strain; TISSUE=Embryo;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Singapore local strain; TISSUE=Embryo;
RG   NIH MGC Project;
RL   Submitted (FEB-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC090521; AAH90521.1; -; mRNA.
DR   InterPro; IPR000074; ApoA1_A4_E.
DR   InterPro; IPR009090; D_amino_pept_C.
FT   NON_TER       1      1
SQ   SEQUENCE   207 AA;  22941 MW;  AF1A0ACCF375A28C CRC64;

  Query Match             9.1%;  Score 109;  DB 2;  Length 207;
  Best Local Similarity   25.5%;  Pred. No. 0.077;
  Matches   37;  Conservative   28;  Mismatches   66;  Indels   14;  Gaps    6;

Qy         76 AVFQCAQCHAVLADSVHLAWDLS-RSLGAVVFSRVTNNVVL-EAPFLVGIEGSLKGSTYN 133
              ||| | :|   : ||  |:|  |       ::  |:|:|:|: : ||: |    |    |
Db         54 AVFMCGKCKLPIGDS--LSWAGSDDENNQIMLKRITDNIVVGKEPFVSGTRKELGCLVVN 111

Qy        134 LLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLS 193
              |  |  |   :|    ||   |   |  || : : :  |:: :       |:| :: |::
Db        112 LT-CRGCCSELGVMYISTPKKLDCKRSLFCFNVENIESYVVGSPG-QQMPELDREDKPVT 169

Qy        194 --------EKIAELKEKIVLTHNRL 210
                     ::: |:    |:   ||
Db        170 LEYQDTVHQQMTEITSLAVIIGQRL 194


RESULT 9
Q54HI6_DICDI
ID   Q54HI6_DICDI PRELIMINARY;      PRT;   532 AA.
AC   Q54HI6;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=DDB0188417;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA   Wardroper A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA   Loulseged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA   Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D., Sanders M.,
RA   Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA   Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA   Shaulsky G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA   Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA   Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 0:0-0(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAFI01000207; EAL62738.1; -; Genomic_DNA.
KW   Hypothetical protein.
```

```
SQ   SEQUENCE   532 AA;  60715 MW;  BAF75AB8CD7C1D5C CRC64;

  Query Match             8.9%;  Score 107;  DB 2;  Length 532;
  Best Local Similarity   22.9%;  Pred. No. 0.4;
  Matches   38;  Conservative   29;  Mismatches   75;  Indels   24;  Gaps    5;

Qy         77 VFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKGST----- 131
              :| |  |   : ||  :   :::|    ::  :|| |::       | |  | ||
Db        205 IFSCLSCKTPIGDSTLI---VNKSHECDMYMTLTNTFVMKKLGGSVILGDQKPSTYRGET 261

Qy        132 ---YNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLL-KTKAIVNASEMDI 187
                 |:| :|  |  |:|      |      |  || :    :  | :     |:   : :|
Db        262 DCIYSLFYCSYCNSPLGRKYNKTSQKFKILNNHFVVDIGSITYYSVGGNNNIIVEDKNEI 321

Qy        188 QNVPLS----------EKIAELKEKIVLTHNRLKSLMKILSEVTPD 223
                ||||          .| : :: |||    ::| |:   : :  |:
Db        322 --VPLSFGNDKLTKMHETMMKMNEKINFLEDQLLSVFNSIKKALPE 365


RESULT 10
YCZC_SCHPO
ID   YCZC_SCHPO     STANDARD;      PRT;   155 AA.
AC   Q9P802;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Hypothetical protein C970.12 in chromosome III.
GN   ORFNames=SPCC970.12;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M.H., Lyne R., Stewart A.,
RA   Sgouros J.G., Peat N., Hayles J., Baker S.G., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D.E., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K.D., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K.L., Murphy L.D., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K.M., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M.N., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J.R., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Duesterhoeft A., Fritzc C., Holzer E., Moestl D.,
RA   Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA   Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SIMILARITY: To human C21orf45.
CC   ---------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   ---------------------------------------------------------------------
DR   EMBL; AL031530; CAB72327.1; -; Genomic_DNA.
DR   GeneDB_Spombe; SPCC970.12; -.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0031055; P:chromatin remodeling at centromere; IMP.
DR   GO; GO:0007059; P:chromosome segregation; IMP.
DR   GO; GO:0016575; P:histone deacetylation; IMP.
DR   GO; GO:0031066; P:regulation of histone deacetylation at cent. . .; IMP.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   155 AA;  17874 MW;  A7AEBD0F93760B0B CRC64;

  Query Match             8.8%;  Score 105;  DB 1;  Length 155;
  Best Local Similarity   30.1%;  Pred. No. 0.13;
  Matches   41;  Conservative   18;  Mismatches   61;  Indels   16;  Gaps    6;

Qy         71 QPERCAVFQCAQCHAVLADSVHLAWDLSRS--LGAVVFSRVTNNVVLEAPFLVGIEGSLK 128
              ||   :|||| :|  :: ||   || :|     |   :   | |:| :|  |    :|
Db         19 QP---SVFQCKKCFQIVGDS--NAWVISHREYLSFTLSDAVENSVRVEDTFKRSDDGL-- 71

Qy        129 GSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQ 188
               |: | |  |   :|    ||    |  :|  :  | ||:  | |  |  ||   :
Db         72 -CVYSELSCTRCNEVIGKVYNSTPIYLDDIRDMYTFSMDKLQAYQLGNKT-VNPEGLTRY 129

Qy        189 NVPLSEKIAELKEKIV 204
                | |     |::| |:
Db        130 QVDL-----EMREDII 140


RESULT 11
Q4Q626_LEIMA
ID   Q4Q626_LEIMA PRELIMINARY;      PRT;  1556 AA.
AC   Q4Q626;
DT   13-SEP-2005 (TrEMBLrel. 31, Created)
DT   13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT   13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE   Hypothetical protein.
GN   ORFNames=LmjF31.2350;
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Friedlin;
RA   Peacock C.S., Murphy L., Ivens A.C., Berriman M., Blackwell J.,
RA   Smith D., Collins M., Fosker N., Harris D., Oliver K., O'Neil S.,
RA   Saunders D., Seeger K., Warren T., Rajandream M., and Barrell B.G.;
RL   Submitted (JUN-2005) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; CT005268; CAJ08424.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1556 AA;  170195 MW;  F6FF7CDDFE8DCA59 CRC64;

  Query Match             7.8%;  Score 94;  DB 2;  Length 1556;
  Best Local Similarity   24.3%;  Pred. No. 27;
  Matches   45;  Conservative   22;  Mismatches   50;  Indels   68;  Gaps    9;

Qy         22 GGTERAIDQASFTTSME------WDTQVVKGS----------SPLGPAGLGAEEPAAGPQ 65
              || |      | |:  :|      :|  | : |              | |  | | |: |||
Db        833 GGGELMQTDAGFSIEVESPIAFLYDDDVHRASQDVGQKRAREEPDGVGGAGEEDAAAG-- 890

Qy         66 LPSWLQPERCAVFQCAQCHAVLADSV-------HLAWDLSRSLGAVVFSRVTNNVVLEAP 118
                  | |: :  |     | |:|        |:   :   :      : :  :|:|:||
Db        891 -------ARAAIGE--QDFGFLNDTVAAPSTYSHMLCRVLELIRTARLNLILLDVLLQAP 941

Qy        119 FLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVC----YLL 174
              |         : |           | :|||             |||||:||  |    :||
Db        942 VL---------TRY-----------VWYHLYK----------HFCLSADKASCVIGMWLL 971

Qy        175 KTKAI 179
              :  |:
Db        972 RNLAV 976


RESULT 12
Q8S611_ORYSA
ID   Q8S611_ORYSA PRELIMINARY;      PRT;   361 AA.
AC   Q8S611;
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE   Putative gypsy-type retrotransposon protein.
GN   ORFNames=OSJNBa0096E22.5;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC099400; AAL91599.1; -; Genomic_DNA.
DR   EMBL; AE017047; AAP51763.1; -; Genomic_DNA.
DR   Gramene; Q8S611; -.
SQ   SEQUENCE   361 AA;  38804 MW;  F883BB3E8B8FFF45 CRC64;

  Query Match             7.7%;  Score 92;  DB 2;  Length 361;
  Best Local Similarity   25.3%;  Pred. No. 6.4;
  Matches   61;  Conservative   28;  Mismatches   92;  Indels   60;  Gaps   14;

Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPA 61
              |  ||  |          |  |||  |    : |     |  || ||    |:|  | :|
Db         49 APTPLPFRRAVRAKAAQDGSGGTSSASPAVAST-----DVVVVPGSREATPSG-PASDPV 102

Qy         62 AGPQLP----SW--LQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVL 115
              ||   |    ||  || |   : :      |:   :  |  |   :     | : ||
Db        103 AGRGSPAAVLSWEELQVEMGRLLEAGA--RVIGREIAEARGLEHRM-----SELGNN--- 152

Qy        116 EAPFLVGIEGSLKGSTYNLL--FCGSCGI-------PVGFHLYSTHAALAAL-----RGH 161
                | | |||:  ||  |    | |||        |  | | |: | |||         |
Db        153 ----LSEIRGSLR-VTYTGLHQLAGKCGIKSTIPANPDEFSLTSSLAELAAAMEEIPSKH 207

Qy        162 FCLSSDKM-------VCYLLKTKAIVNASEMDIQNV-----------PLSEKIAELKEKI 203
                ::|        |::|    : :  |:|:: :            : |:: :| : :
Db        208 AARIGEEMSNRIYIGACHILACVRLAH-PELDLREILDQGEASDARKDVMEEVGDLGKSV 266

Qy        204 V 204
              :
Db        267 L 267


RESULT 13
Q9CH87_LACLA
ID   Q9CH87_LACLA PRELIMINARY;      PRT;   695 AA.
AC   Q9CH87;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Copper-potassium transporting ATPase B.
GN   Name=copB; OrderedLocusNames=LL0851;
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr.GR-1697R;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
DR   EMBL; AE006319; AAK04949.1; -; Genomic_DNA.
DR   PIR; C86731; C86731.
DR   GO; GO:0016021; C:integral to membrane; IEA.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0004008; F:copper-exporting ATPase activity; IEA.
DR   GO; GO:0016820; F:hydrolase activity, acting on acid anhydrid. . .; IEA.
DR   GO; GO:0046872; F:metal ion binding; IEA.
DR   GO; GO:0046873; F:metal ion transporter activity; IEA.
DR   GO; GO:0008152; P:metabolism; IEA.
DR   GO; GO:0030001; P:metal ion transport; IEA.
DR   GO; GO:0015992; P:proton transport; IEA.
DR   InterPro; IPR006403; ATPase-IB1_Cu.
DR   InterPro; IPR006416; ATPase-IB_hvy.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR005834; Dehal_like_hydro.
DR   InterPro; IPR008250; E1-E2_ATPase_reg.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   TIGRFAMs; TIGR01511; ATPase-IB1_Cu; 1.
DR   TIGRFAMs; TIGR01525; ATPase-IB_hvy; 1.
DR   TIGRFAMs; TIGR01494; ATPase_P-type; 2.
DR   PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   695 AA;  75981 MW;  1B4947C3C2A0FA0F CRC64;

  Query Match             7.7%;  Score 92;  DB 2;  Length 695;
  Best Local Similarity   25.3%;  Pred. No. 15;
  Matches   38;  Conservative   28;  Mismatches   50;  Indels   34;  Gaps    7;

Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
              || ::|           ||  | :| ||    ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI-------------QNVPLSE-K 195
              |          |    :: :|:: |||: | :   |:             ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
              : :|   :|  : :: :   ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241


RESULT 14
Q69KE0_ORYSA
ID   Q69KE0_ORYSA PRELIMINARY;      PRT;   628 AA.
AC   Q69KE0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Myosin heavy chain-like.
GN   Name=P0023E10.17;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Sasaki T., Matsumoto T., Hattori M., Sakaki Y., Katayose Y.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 9, PAC
RT   clone:P0023E10.";
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005934; BAD36600.1; -; Genomic_DNA.
DR   Gramene; Q69KE0; -.
DR   InterPro; IPR005829; Sug_transporter.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
SQ   SEQUENCE   628 AA;  68281 MW;  3230B108C6061027 CRC64;
```

```
  Query Match             7.6%;  Score 91;  DB 2;  Length 628;
  Best Local Similarity   23.7%;  Pred. No. 16;
  Matches   66;  Conservative   34;  Mismatches   82;  Indels   96;  Gaps   17;

Qy          8 HRSRCATPPRG----------DFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGA 57
              :| :  ||              |  |||  |    : |     |  || ||    |:|| |
Db        141 YRRKAPTPSPSRRVARAKAAQDGSGGTSSASPAVAST-----DVVVVPGSREAAPSGL-A 194

Qy         58 EEPAAGPQLP----SW--LQPERCAVFQC---------AQCHAVLADSVH-----LAWDL 97
               :  |||  |    :|  || |   : :          |:  |  | | :     || ||
Db        195 SDLVAGPGSPAAVLTWEELQVEMGRLLEAGARGIGREIAEARAAAASSANERADRLAHDL 254

Qy         98 S------RSLGAVV------------FSRVTNNVVLEAPFLVGIEGSLKGSTYNLL--F 136
              :      : :  :|            | : ||       |  |  ||:  ||  |
Db        255 AEAREDLKKMRELVAGNERQWQGLEHRMSELENN-------LSEIRDSLR-VTYTGLHQL 306

Qy        137 CGSCGI-------PVGFHLYSTHAALAAL-----RGHFCLSSDKM-------VCYLLKTK 177
              | |||       |  | | |: | |||       |   : ::        |::|
Db        307 AGECGIKSTIPANPDEFSLTSSLAELAAAMEEIPSKHAARTREETSNGIYTGACHVLACL 366

Qy        178 AIVNASEMDIQNV----------PLSEKIAELKEKIV 204
              :|:  |:|:: :          : |:: :| | ::
Db        367 RLVH-PELDLREILDQGAASDARKDVMEEVGDLGESVL 403


RESULT 15
O83693_TREPA
ID   O83693_TREPA PRELIMINARY;      PRT;   597 AA.
AC   O83693;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Phosphoribosylglycinamide formyltransferase, putative.
GN   OrderedLocusNames=TP0695;
OS   Treponema pallidum.
OC   Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Nichols;
RX   MEDLINE=98332770; PubMed=9665876; DOI=10.1126/science.281.5375.375;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R.J., Gwinn M.L., Hickey E.K., Clayton R.A., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S.L., Peterson J.D.,
RA   Khalak H.G., Richardson D.L., Howell J.K., Chidambaram M.,
RA   Utterback T.R., McDonald L.A., Artiach P., Bowman C., Cotton M.D.,
RA   Fujii C., Garland S.A., Hatch B., Horst K., Roberts K.M., Sandusky M.,
RA   Weidman J.F., Smith H.O., Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
DR   EMBL; AE001243; AAC65662.1; -; Genomic_DNA.
DR   PIR; D71293; D71293.
DR   TIGR; TP0695; -.
DR   GO; GO:0016874; F:ligase activity; IEA.
DR   InterPro; IPR003135; ATP-grasp.
DR   InterPro; IPR011761; ATP_GRASP.
DR   Pfam; PF02222; ATP-grasp; 1.
DR   PROSITE; PS50975; ATP_GRASP; 1.
KW   Complete proteome.
SQ   SEQUENCE   597 AA;  63326 MW;  7B775117C9BFD6A6 CRC64;

  Query Match             7.6%;  Score 90.5;  DB 2;  Length 597;
  Best Local Similarity   28.0%;  Pred. No. 17;
  Matches   51;  Conservative   19;  Mismatches   47;  Indels   65;  Gaps   12;

Qy         12 CATPPRGDFCGGTERAID----QASFT--------------TSMEWDTQVVKGSSPLGPA 53
              || |  |      |: | |    :| ||              |  |||       | | |
Db         90 CALP--GHRLEATKNATDKTRMRACFTRARLRCPRFTFLEPDSFAWDT-------PPGHA 140

Qy         54 GLGAEEPAAGPQLPSWLQP-ERCAVFQC--AQCHAVLADSVHLAWDLSRSLGAVVFSRVT 110
               | :   :||   |  ::| :      |  |||   | ::  :|   |||

Db        141 RLCSHLHSAGLSFPLVVKPTDNMGARGCTLAQCKDTLINACAVARQFSRS---------- 190

Qy        111 NNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMV 170
                 |::|  |:|| | ||:|    |:| |:        || |  |||          |: :
Db        191 GRVIIE-EFIVGREFSLEG----LIFDGT--------LYVT--ALA----------DRHI 225

Qy        171 CY 172
               |:
Db        226 CF 227


Search completed: November 18, 2005, 20:35:05
Job time : 237 secs
```